

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:44:16 ; Search time 44 Seconds

(Without alignments)
345,210 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906
Sequence: 1 MROLNFCSELOSPQOSYLO.....PPQOPQPPVPOQASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:*
2: PIR.*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	830.5	91.7	260	2 S18350	seed storage prote
2	752	83.0	310	2 T06211	C-hordein precursor
3	580	64.0	347	2 T05277	probable hordein C
4	566	62.5	357	2 S18236	omega secalin prec
5	562.5	62.1	357	2 S18235	gamma secalin prec
6	412	45.5	327	2 JS0402	gamma-gliadin prec
7	384.5	42.4	302	2 JA0153	gamma-gliadin prec
8	347	38.3	251	2 PS0094	gamma-gliadin prec
9	332	36.6	291	1 EEMT	gamma-gliadin B pr
10	310.5	34.3	105	2 S07189	gamma-gliadin (clone p
11	302	33.3	305	2 S08312	C-hordein (clone p
12	289.5	32.0	307	2 S10015	gamma-hordein 1 pr
13	279.5	30.8	509	2 A48048	alpha-beta-gliadin
14	272.5	30.1	72	2 A25677	egg envelope prote
15	271.5	30.0	293	2 S07365	C-hordein (pc horl
16	268.5	29.6	326	2 D22364	hordein B1 - barley
17	263.5	29.1	297	2 T06500	alpha-beta-gliadin
18	262.5	29.0	320	2 S20519	alpha-beta-gliadin
19	260.5	28.8	320	2 E22364	hordein B precursor
20	256.5	28.3	286	2 T05718	alpha-beta-gliadin
21	255.5	28.2	21	2 C22364	gamma-hordein 3 -
22	255	28.1	296	2 S07361	alpha-beta-gliadin
23	253.5	28.0	296	2 A27319	alpha-beta-gliadin
24	250	27.6	291	2 T06498	gliadin - wheat
25	244.5	27.0	1343	2 AF0611	cell division prot
26	242.5	26.8	374	2 T05923	glutenin low molec
27	242	26.7	282	2 T06504	alpha-beta-gliadin
28	240	26.5	264	2 S07975	B3-hordein (clone
29	239	26.4	313	2 S07924	alpha-beta-gliadin

30	238	26.3	356	2 S01992	glutenin low molec
31	238	26.3	359	2 T06982	glutenin low molec
32	237	26.2	288	2 T06282	alpha-gliadin prec
33	236.5	26.1	319	2 A22364	alpha-beta-gliadin
34	236	26.0	286	2 S07923	glutenin low molec
35	235.5	26.0	276	2 S57636	glutenin low molec
36	234	25.8	286	1 EEMT	alpha-beta-gliadin
37	234	25.8	292	2 B22364	alpha-beta-gliadin
38	234	25.8	1342	2 G90750	cell division prot
39	234	25.8	1342	2 E85614	cell division prot
40	230.5	25.4	261	2 S57655	glutenin low molec
41	219	24.2	271	2 T04474	B1 hordein - barley
42	218.5	24.1	68	2 B25677	C-hordein (pc-919)
43	217	24.0	1329	2 A64828	cell division prot
44	211.5	23.3	307	2 S04325	glutenin low molec
45	208.5	23.0	462	1 LDD07	annexin VII - slim

ALIGNMENTS

RESULT 1
S18350
seed storage protein - barley
C/Species: Hordeum vulgare (barley)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C/Accession: S18350
R:Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron Mills, V.
Plant Mol. Biol. 17, 1217-1231, 1991
A>Title: Amber codon suppression: the in vivo and in vitro analysis of two C-hordein
A:Reference number: S18350; MUID:92032786; PMID:1932695
A:Accession: S18350
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <EMT>
A:Cross-references: EMBL:X60037; NID:q19000; PIDN:CAA42642.1; PID:q19001
C:Superfamily: gliadin

Query Match
Best Local Similarity 91.7% Score 830.5; DB 2; Length 260;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 ROLNFCSELOSPQOSYLOPPYQPNFYLPQKPPVQOQPPFPYLPQYFPELPQYQIP 61
DB 21 ROLNFCSELOSPQOSYLOPPYQPNFYLPQKPPVQOQPPFPYLPQYFPELPQYQIP 80
QY 62 TPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 121
DB 81 TPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 140
QY 122 QITFQPPQOSYPPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 150
DB 141 QITFQPPQOSYPPVQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 168

RESULT 2
T06211
C-hordein precursor - barley
C/Species: Hordeum vulgare (barley)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C/Accession: T06211; A03355
R:Salinova, O.V.; Mekhedov, S.L.; Zheludin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.
Genetika 29, 1070-1079, 1993
A>Title: Nucleotide sequence of the barley C-hordein gene.
A:Reference number: Z15537; MUID:93380629; PMID:8396543
A:Accession: T06211
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-310 <SAT>
A:Cross-references: EMBL:S6638; NID:q442523; PIDN:AAB28161.1; PID:q442524
R:Shewry, P.R.; Autran, J.C.; Nimmo, C.C.; Lew, E.J.L.; Kasarda, D.D.
Nature 286, 520-522, 1980
A:Reference number: A93228

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:57:50 ; Search time 150.754 Seconds
(without alignments)
4178.078 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 1 atgagagcaactaaaccttg.....gcatagtgagatgcttag 477
Sequence:

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106.2	22.3	1635	10	US-09-864-761-20241
2	106.2	22.3	1973	10	US-09-864-761-3471
3	103	21.6	15720	9	US-10-025-380-1058
4	103	21.6	15720	10	US-09-922-217-1058
5	103	21.6	15720	10	US-09-833-263-1058
6	100.2	21.0	1075	10	US-09-864-761-19241
7	100.2	21.0	1403	10	US-09-864-761-2513
8	94.4	19.8	867	10	US-09-216-393-340
9	94.4	19.8	867	10	US-09-216-393-342
10	94.4	19.8	1397	10	US-09-216-393-343
11	94.4	19.8	1397	10	US-09-216-393-345
12	93.6	19.6	1236	9	US-10-077-584-3
13	92.6	19.4	446	10	US-09-864-761-20699
14	91.2	19.1	830	10	US-09-864-761-19531
15	77.4	16.2	479	10	US-09-864-761-1768
16	77.2	16.2	401	10	US-09-864-761-3936
17	77	16.1	155074	9	US-10-026-188-6
18	70.2	14.7	600	10	US-09-864-761-6690
19	69.4	14.5	887	10	US-09-864-761-21484

C 20	69.4	14.5	1981	10	US-09-864-761-4745	Sequence 4745, Ap
C 21	67.8	14.2	2901	10	US-09-801-368-371	Sequence 371, App
C 22	67.4	14.1	520	9	US-10-184-644-332	Sequence 332, App
C 23	67.4	14.1	520	9	US-10-184-634-332	Sequence 332, App
C 24	66.6	14.0	493	10	US-09-864-761-5863	Sequence 5863, Ap
C 25	64.6	13.5	1322	9	US-09-789-054A-9	Sequence 9, App11
C 26	62.8	13.2	574	10	US-09-864-761-228	Sequence 228, App
C 27	62.8	13.2	669	10	US-09-864-761-17051	Sequence 17051, A
C 28	62.4	13.1	573	10	US-09-864-761-5592	Sequence 5592, App
C 29	61.8	13.0	350	10	US-09-864-761-23424	Sequence 23424, A
C 30	61.8	13.0	522	10	US-09-864-761-19900	Sequence 19900, A
C 31	60.8	12.7	549	9	US-10-255-536-208	Sequence 208, App
C 32	60.6	12.7	2656	10	US-09-842-552-82	Sequence 82, App1
C 33	60.4	12.7	968	10	US-09-864-761-9629	Sequence 9629, Ap
C 34	60	12.6	925	9	US-09-735-056-1	Sequence 1, App11
C 35	59.2	12.4	470	10	US-09-864-761-3121	Sequence 3121, Ap
C 36	57.8	12.1	497	10	US-09-864-761-22361	Sequence 22361, A
C 37	57.8	12.1	1818	10	US-09-801-368-427	Sequence 427, App
C 38	57.4	12.0	1134	10	US-09-737-178-111	Sequence 111, App
C 39	57.4	12.0	1137	10	US-09-737-178-104	Sequence 104, App
C 40	57	11.9	537	10	US-09-864-761-8390	Sequence 8390, App
C 41	55.8	11.7	1089	10	US-09-925-302-301	Sequence 301, App
C 42	55.6	11.7	3150	9	US-09-938-842A-243	Sequence 243, App
C 43	55.6	11.7	4104	10	US-09-801-368-107	Sequence 107, App
C 44	55.4	11.6	476	10	US-09-864-761-6	Sequence 6, App11
C 45	55.2	11.6	768	9	US-09-938-842A-812	Sequence 812, App

ALIGNMENTS

RESULT 1
US-09-864-761-20241
: Sequence 20241, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Hanzel, David R.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
: FILE REFERENCE: Aecmeca-X-1
: CURRENT APPLICATION NUMBER: US/09/864, 761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180, 312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207, 456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632, 366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263, 6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236, 359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-20241

```

```

Query Match      22.3%; Score 106.2; DB 10; Length 1635;
Best Local Similarity 53.2%; Pred. No. 1.1e-19;
Matches 225; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 34 CAATACCAACCAATCATATCTGCGGAGCATATGCAAAACCAATATACCGCAA 93
DB 501 CCATACCTACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 560
QY 94 AAACATTTCCAGTGCAGCAACCGTTTCACACGCCCAACATATTTCCCTATCA 153
DB 561 CCACCAACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 620
QY 154 GAGGAATTTGTTCCCAATATCAATACCAACCGCCCAACCAACCAACCAATCCC 213
DB 621 CCACCTACTGTCGACCAACCAACCAACCAATCAGTACCAACCAACCAATCTG 680
QY 214 CAACAACCAACCAACCAACCTTTCTGCGCCCAACCAACCAATTTCCCTGCAACCA 273
DB 681 CCACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 740
QY 274 CCATTTCCCGAGCCCAAGAACCAATTTCCCAACCAATTTCCCTGCAACCAACCA 333
DB 741 CTACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 800
QY 334 CCATTTCCCGAGCCCAAGAACCAATTTCCCAACCAATTTCCCTGCAACCAACCA 393
DB 801 CCATACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 860
QY 394 TACCTGTGCAACCTCAACAGGCAATTTCTTAACCACTCAACCAAGTCCCAAGCT 453
DB 861 CCACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 920
QY 454 TCA 456
DB 921 TCA 923

```

RESULT 2

```

US-09-864-761-3471
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
; LENGTH: 1973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471

```

```

Query Match      22.3%; Score 106.2; DB 10; Length 1973;
Best Local Similarity 53.2%; Pred. No. 1.2e-19;
Matches 225; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 34 CAATACCAACCAATCATATCTGCGGAGCATATGCAAAACCAATATACCGCAA 93
DB 655 CCATACCTACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 714
QY 94 AAACATTTCCAGTGCAGCAACCGTTTCACACGCCCAACATATTTCCCTATCA 153
DB 715 CCACCAACCAACCAACCAACCAATCAGTACCAACCAACCAATCA 774

```


OTHER INFORMATION: NT HIT: AL163201.2, EVALUATE 2.00e-19
US-09-864-761-19241

Query Match 21.0%, Score 100.2; DB 10; Length 1075;

Best Local Similarity 52.2%; Pred. No. 4.1e-18;

Matches 222; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 32 TGCATACCAACCAATCATATCTGCGCGCATATCCCAAAACCAATATCTACCG 91
DB 943 TACCACTACCAACCAATCATATCTGCGCGCATATCCCAAAACCAATATCTACCG 884
QY 92 AAAAACCATTTCAGTGAGCAACCGTTTCACACACCCCAACAAATATTTCCCTATCTAC 151
DB 883 CACACACCACTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 824
QY 152 CAGAGGAATTTTTCCTCAATATATCAATACCAACCCCTTCAACCAACCAACCAATTC 211
DB 823 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 764
QY 212 CCAACCAACCAACCAACCACTCTCTGCGCGCAACCAACCACTCTGCGCGCAACCAAC 271
DB 763 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 704
QY 272 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCGCAACCAAC 331
DB 703 CATGACCAACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 644
QY 332 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCGCAACCAAC 391
DB 643 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 584
QY 392 CATACCTGTTGCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 451
DB 583 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 524
QY 452 CTTC A 456
DB 523 CATCA 519

RESULT 7
US-09-864-761-2513/c
Sequence 2513, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIORITY FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263,6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2513
LENGTH: 1403
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL078472.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

Query Match 21.0%, Score 100.2; DB 10; Length 1403;
Best Local Similarity 52.2%; Pred. No. 4.6e-18;
Matches 222; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 32 TGCATACCAACCAATCATATCTGCGCGCATATCCCAAAACCAATATCTACCG 91
DB 1293 TACCACTACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTAC 1234
QY 92 AAAAACCATTTCAGTGAGCAACCGTTTCACACACCCCAACAAATATTTCCCTATCTAC 151
DB 1233 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1174
QY 152 CAGAGGAATTTTCCCAATATATCAATACCAACCCCTTCAACCAACCAACCAATTC 211
DB 1173 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1114
QY 212 CCAACCAACCAACCAACCACTCTCTGCGCGCAACCAACCAATTTCCCTGCGCAACCAAC 271
DB 1113 CACCATGACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 1054
QY 272 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCGCAACCAAC 331
DB 1053 CATGACCAACCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 994
QY 332 AACCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCGCAACCAAC 391
DB 993 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 934
QY 392 CATACCTGTTGCAACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 451
DB 933 CACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTAC 874
QY 452 CTTC A 456
DB 873 CATCA 869

RESULT 8
US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
US-09-216-393-340

Query Match 19.8%; Score 94.4; DB 10; Length 867;
Best Local Similarity 50.7%; Pred. No. 1.5e-16;
Matches 227; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 3 GAGGCACTAAACCTTGGCAGCAAGATTGCAATCACCACACATATATCTGCGCGCA 62
DB 388 GAGGAGGCTGCGTTATTAGGCGCAAGAGGTCAAGAGGGAAGGAAAAAGCTACACTACA 447
QY 63 GCCATATCCCAAAACCCATATCTACCGCAAAACCATTTCCAGTGCAGCAACGGTTTCA 122
DB 448 ACCAGTTTCATCCACAGTAGACAGTACAGACGACGATCAATACACTACTACTAC 507
QY 123 CACACCCCAACAAATATTTCCCTATCTACAGAGAAATTTGTTCCCAATATCAAAATACC 182
DB 508 ACCACCACTAGACTACTACTACACTACGACACCAACACACACACACACACACACACA 567
QY 183 AACCCCTTACACCAACCAACCAATTTCCCAACCAACCAACCAACCAACCAACCAAC 242
DB 568 ACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 627
QY 243 CCACACACCAATTTCCCTGGCAGCAACCAACCAATTTCCCGAGCCCAAGAACCAATTC 302
DB 628 ACAACACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 687
QY 303 CCACACCAATTTCCCTGGCAGCAACCAACCAATTTCCCGAGCCCAAGAACCAATTC 362
DB 688 ACAACACTACAGACTACCAACCAACTACTACCACTACCAACCAACCAACCAACCAACCA 747
QY 363 ACAATAATTTTCCAGCAACCAACCAATCTACCTGTGCAACCTCAACCAACCAATTC 422
DB 748 CCACACACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 807
QY 423 TCACACACCTCAACCAAGTCCCAACCA 450
DB 808 ACAACTACACAGACTACGACCACTCTA 835

RESULT 9
US-09-216-393-342
; Sequence 342, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18

EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 342
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
US-09-216-393-342

Query Match 19.8%; Score 94.4; DB 10; Length 867;
Best Local Similarity 50.7%; Pred. No. 1.5e-16;
Matches 227; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 3 GAGGCACTAAACCTTGGCAGCAAGATTGCAATCACCACACATATCTGCGCGCA 62
DB 388 GAGGAGGCTGCGTTATTAGGCGCAAGAGGTCAAGAGGGAAGGAAAAAGCTACACTACA 447
QY 63 GCCATATCCCAAAACCCATATCTACCGCAAAACCATTTCCAGTGCAGCAACGGTTTCA 122
DB 448 ACCAGTTTCATCCACAGTAGACAGTACAGACGACGATCAATACACTACTACTAC 507
QY 123 CACACCCCAACAAATATTTCCCTATCTACAGAGAAATTTGTTCCCAATATCAAAATACC 182
DB 508 ACCACCACTAGACTACTACTACTACGACACCAACCAACCAACCAACCAACCAACCAACA 567
QY 183 AACCCCTTACACCAACCAACCAATTTCCCAACCAACCAACCAACCAACCAACCAAC 242
DB 568 ACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 627
QY 243 CCACACACCAATTTCCCTGGCAGCAACCAACCAATTTCCCGAGCCCAAGAACCAATTC 302
DB 628 ACAACACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 687
QY 303 CCACACCAATTTCCCTGGCAGCAACCAACCAATTTCCCGAGCCCAAGAACCAATTC 362
DB 688 ACAACACTACAGACTACCAACCACTACTACTACCACTACCAACCAACCAACCAACCAAC 747
QY 363 ACAATAATTTTCCAGCAACCAACCAATCTACCTGTGCAACCTCAACCAACCAATTC 422
DB 748 CCACACACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 807
QY 423 TCACACACCTCAACCAAGTCCCAACCA 450
DB 808 ACAACTACACAGACTACGACCACTCTA 835

RESULT 10
US-09-216-393-343
; Sequence 343, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 343
; LENGTH: 1397
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (238)..(1104)
US-09-216-393-343

Query Match 19.8%; Score 94.4; DB 10; Length 1397;
Best Local Similarity 50.7%; Pred. No. 1.9e-16;

Matches 227; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

OY	3	GAAGCAACTTAACCCCTTGAGCAAGAACTTCATTCACCAACCAATCATATTCGGCGA	62
Db	625	GABGAGGCTGGCTTATTATGCGCAAGAGGTGCAAGAGAGGAAAAAGACTACACTACA	684
OY	63	GCCATATCCACAAACCCATATCTACGGCAAAAACATTCGAGTGACAAACCGTTTCA	122
Db	685	ACCAGTTCTCCACAAGTAGCAGTAGTACAAGCACACGATCACTACACTTACACT	744
OY	123	CATCCCCCAACATATTTTCCCTATCTACGACAGAGAAATTTGTTTCCCATATTCATAC	182
Db	745	ACCAACCCATGAGACTACTACTACACTTACGACACCAACAACAATCAACAACACACACA	804
OY	183	AACCCCTCTACAGCAACAACACATTTTCCCAACAACCAACCAACCTTTCTCGGCC	242
Db	805	ACTACACCAACAACAAGCAACCAACCAACCACTACACCAACAACAAGCAACAACCC	864
OY	243	CCAGCAACACTTCCCTGGCAACCAACAACAATTTCCGAGCCCCAGAACCAATTC	302
Db	865	ACACACTACTACCAACAACAAGCAACAACCAAGCCCAACTACAAAGCAATCTAG	924
OY	303	CCAGCAACACTTCCCTGGCAACCAACAACAATTTCCGAGCCCCAGAACCAATTC	362
Db	925	ACAACCACTACAGACTTACCAACACTTACTCTACACCACTTCAACACAGCAACCAAGGA	984
OY	363	ACAAATTAATTTTCGAGCAACCCCAACAATCATCTACCTGTGCAACCTCAACGCAATTC	422
Db	985	CCAACTACTACAGCAACAACCAAGCAACCAACCACTACGAAGCAACAAGCACTAG	1044
OY	423	TCAAGAACTCAACAAGTCCCAACA	450
Db	1045	ACAAGTACAACGACTTACGACACCACTTA	1072

```

RESULT 11
US-09-216-393-345/c
: Sequence 345, Application US/09216393
: Patent No. US2001001447A1
: GENERAL INFORMATION:
: APPLICANT: Milhausen, Michael James
: TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
: FILE OF INVENTION: USES THEREOF
: FILE REFERENCE: TX-1-C2
: CURRENT APPLICATION NUMBER: US/09/216,393
: CURRENT FILING DATE: 1998.12.18
: EARLIER APPLICATION NUMBER: 08/994,825
: EARLIER FILING DATE: 1997-12-19
: NUMBER OF SEQ. ID NOS.: 364
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO. 345
: LENGTH: 1397
: TYPE: DNA
: ORGANISM: Toxoplasma gondii
US-09-216-393-345

```

Query Match	19.88;	Score 94.4;	DB 10;	Length 1337;
Best Local Similarity	50.7%;	Pred. No. 1.9e-16;		
Matches 227; Conservative	0;	Mismatches 221;	Indels 0;	Gaps 0

Qy	5	GAGGCACAAATTAACCCCTTGAGCCAGAGAGTGTGCATACACAAACAAATCATATCTGGCCGA	62
Db	773	GAGAGAGCTCCGCTTCTATTAGGCCAAGAGAGTCAAGAGAGGAGAAAAAGCATACACTACCA	714
Qy	63	GCCATATTCACAAATACCCATATCTACCCGCAAAAACATTTGCATGCGACGACCCGTTTCA	122
Db	713	ACCAAGTTTCACCAAGTAGAGAGTACCAAGACACAGACATCACTACCTACTTACCAC	654
Qy	123	CACACCCCAACATTAATTTCCCTATCTACAGAGSAATTTGTTTCCCAATATCAATAC	182
Db	653	ACGACACGCTAGCACTACTACGAAGTACGACACCAACAAACATACAAACACACACA	594
Qy	183	AACCCCTTCAACACCAACACCATTTCCCCCAACACCAACACCACTTTCTCGGCC	242

[illegible]

```

RESULT 12
US-10-077-584-3
: Sequence 3, Application US/10077584
: Publication No. US20030073610A1
: GENERAL INFORMATION:
: APPLICANT: LINDQUIST, SUSAN
: APPLICANT: KROBITSCH, SYLVIA
: APPLICANT: OUTEIRO, TIAGO F.
: TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
: FILE REFERENCE: ARD:36705
: CURRENT APPLICATION NUMBER: US/10/077,584
: CURRENT FILING DATE: 2002-02-15
: PRIOR APPLICATION NUMBER: 60/269,157
: PRIOR FILING DATE: 2001-02-15
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO. 3
: LENGTH: 1236
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(513)
US-10-077-584-3

```

Query Match	19.68;	Score 93.6;	DB 9;	Length 1236;
Best Local Similarity	51.28;	Pred. No. 3e-16;		
Matches 219; Conservative	0;	Mismatches 209;	Indels 0;	Gaps 0;

Oy	22	GCAGCCAGAGCTTCATTCACGACAGCATCTATATCTCGCGAGCATATCCACAAC	79
	11		11111
	47	GCCTCCAAACAGCAGCAACAGCAACACAGCAGCAACAGCAACAGCAACAGCAAC	106
Oy	80	CATATCTACCCCAAAACCATTTCTCGATCGACAGACGCTTTACACACCCCAACATATT	133
	11		11111
Db	107	AACACAGCAACAGCAACCAACAGCAACCAACAGCAACAGCAACAGCAACAGCAAC	166
Oy	140	TCCCGTATCTACGAGAGAAATTTCTTCCCGCATTTATCAATACAGCCGCTACAAAC	199
Db	167	AGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAC	222
Oy	200	AACACCATTTCCCAACACACACACACACTCTTCTCGGCGCCAGAACCATTCGCT	255
	11		11111
Db	227	AGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAC	286
Oy	260	GGCAACCAACAACCAATTTCCCAAGCCCAAGAACCAATTTCCCAACAACATTTCCCT	319
	11		11111
Db	287	AACAGAGAAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAC	346
Oy	320	GGCAACCAACAACCAATTTCCCAAGCCCAAGAACCAATTTCAACAATTAATTTCCAC	373
	11		11111
Db	347	AGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAC	400

```

OY      380 AACCCACATCATACCTGTGACACCTCAGACGCAATTCCTCAACACCTCAACG 439
      11 111 11 111 111 111 11 111 11 111
Db      407 CTCCTCAGACGACGACGCTGTGCTGCTCAGCAGACACCTGCTCAGCCTCAGC
      440 TCCCCCAA 447
OY      467 CTCCTCA 474
Db      467 CTCCTCA 474

RESULT 13
US-09-864-761-20699/c
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US/60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US/09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US/60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US/60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US/09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US/09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007249.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13

```

```

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
; OTHER INFORMATION: NT HIT: AL161539.2, EVALU = 3.70e+00
US-09-864-761-20699

Query Match      19.4%; Score 92.6; DB 10; Length 446;
Best Local Similarity 51.9%; Pred. No. 3.5e-16;
Matches 209; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

OY      38  CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 97
      111111 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      417 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 358
      111111 11 11 11 11 11 11 11 11 11 11 11 11 11
OY      98  CATTTCAGAGCAGCAACCGTTTACACACCCCAACCAATTTCCCTATTCAGCAGG 157
      11 111 111 111 11 11 11 11 11 11 11 11 11 11
Db      357 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 298
      11 111 111 11 11 11 11 11 11 11 11 11 11
OY      158 AATTGTTCCCAATATCAATATCAACAGCCCTTACCAACCAACCAATTTCCCGCCAGC 217
      11 11 11 11 11 11 11 11 11 11 11 11 11
Db      297 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 238
      11 11 11 11 11 11 11 11 11 11 11 11
OY      218 AACCAACAAACCTTTCTGCGCCCAACCAACCAATTTCCCTGCAACCAACCAACCAT 277
      11111 11 11 11 11 11 11 11 11 11 11 11 11
Db      237 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 178
      11 11 11 11 11 11 11 11 11 11 11 11
OY      278 TTCCTCCAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCAACCAACCAACCAT 337
      11 11 11 11 11 11 11 11 11 11 11 11
Db      177 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 118
      11 11 11 11 11 11 11 11 11 11 11 11
OY      338 TTCCTCCAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCAACCAACCAACCAT 397
      11 11 11 11 11 11 11 11 11 11 11 11
Db      117 CACCTCCAGCCCAAGAACCAATTTCCCAACCAACCAATTTCCCTGCAACCAACCAACCAT 58
      11 11 11 11 11 11 11 11 11 11 11 11
OY      398 CTCGTGACCTTCACAGCAGCATTTCTCAACACCTCAACCAT 440
      11 11 11 11 11 11 11 11 11 11 11 11
Db      57 CACCAACATCATATTCGCGCAGCATATCCACAAACCATTTCTACCGCAAAAC 15
      11 11 11 11 11 11 11 11 11 11 11 11

RESULT 14
US-09-864-761-19531/c
; Sequence 19531, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US/60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US/09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US/60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19531
; LENGTH: 830
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO 283851.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: SWISSPROT HIT: Q28614, EVALUATE 2.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: Z45997.1, EVALUATE 7.90e-01
US-09-864-761-19531
Query Match          19.1%; Score 91.2; DB 10: Length 830;
Best Local Similarity 51.2%; Pred. No. 11e-15;
Matches 213; Conservative 0; Mismatches 203; Indels 0; Gaps 0;
QY 34 CAATACCACCAACATCATATCTGCGCCAGCCATATCCAGAAAACCATATCTACCCGAA 93
Db 579 CCATCCACCATCATCATCATCCACCCACCATTAACCTCCACACACCATTAACCTGCA 520
QY 94 AACCATTTCCAGTGCAGCAACGCTTTCACACACCCCAACAAATATTTCCCTATCTACCA 153
Db 519 TCACCCACCATCATCATCATCTCTATTCACACACACCATCCACCATCAACCA 460
QY 154 GAGGAATTTGTTCCCAATATCAAAATACCAACCCCTTACACACCAACCATTTGCC 213
Db 459 CCACACACCATCATCATCACTCCATCCACACACCATCAACCATTAACCTCATCA 400
QY 214 CAACACACCAACAACCTCTTCTCGGCCCAACCAACCATTTCCCTGCAACCAACAA 273
Db 399 CCACACACCATCATCACTCCATCCATCCACACCAACCATTAACCTCGCCACCA 340
QY 274 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCATTTCCCTGCAACCAACAA 333
Db 339 CCACACCATTAACCTCCACCAACCAACCAACCAACCATTAACCTCATCAACCA 280
QY 334 CCATTTCCCAAGCCCAAGAACCAATTTCCCAACCAACCATTTCCCAACCAACCAACAA 393
Db 279 CCACACCATTAACCTCCATCCATCCACCAACCAACCTTCACACCATTAACCTCATCA 220
QY 394 TACCCTGTGCAACTCAACGCAATTTCTCAACCAACCTCAACCAAGTCCCAACA 449
Db 219 CCACACCATTAACCTCATCACTCTCACTCAACCAACCAACCAACCAACCAACCA 164
RESULT 15
```

```

US-09-864-761-768/c
; Sequence 768, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica X-1
; CURRENT FILING DATE: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 768
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO 283851.17
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
US-09-864-761-768
Query Match          16.2%; Score 77.4; DB 10: Length 479;
Best Local Similarity 51.6%; Pred. No. 6e-12;
Matches 177; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
```

[illegible]

Search completed: May 25, 2003, 20:22:00
Job time : 152.754 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:52:25 ; Search time 1662.16 Seconds

(without alignments)
4647.728 Million cell updates/sec

Title: US-09-743-533-18

Perfect score: 477
Sequence: 1 atgaagcaataaaccttgcg.....gcatacgagatgctcag 477

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8057743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.*
1: em_estbda.*
2: em_estbhm.*
3: em_estlin.*
4: em_estlm.*
5: em_estm.*
6: em_estpl.*
7: em_estro.*
8: em_hlc.*
9: gb_estl.*
10: gb_est2.*
11: gb_hlc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vtl.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_tod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	321.6	67.4	567	14	BQ755360 EBbed07_SQ
2	319	66.9	585	10	BE195280 HVSMEH008
3	312.4	60.5	507	10	BE601687 HVSMEH009
4	287.4	60.3	936	12	BG365870 HVSMEI000
5	246.8	51.7	515	9	AJ435523 AJ435523
6	246.2	51.6	420	9	AJ485191 AJ485191

C	7	245.8	51.5	360	9	AJ484592	AJ484592
C	8	236.8	49.6	300	9	AJ484591	AJ484591
	9	224.2	47.0	940	14	BQ609718	BQ609718
	10	219	45.9	595	14	BQ472381	BQ472381
	11	218.6	45.8	1008	13	B1947525	B1947525
	12	218	45.7	560	14	BQ245760	BQ245760
	13	217.8	45.7	611	14	BQ472392	BQ472392
	14	215	45.1	854	13	B1951055	B1951055
	15	213.8	44.8	706	14	BQ804665	BQ804665
	16	211	44.2	520	14	BQ472916	BQ472916
	17	197.8	41.5	494	12	BQ366682	BQ366682
	18	185.6	38.9	822	12	BQ366988	BQ366988
	19	182.6	38.3	762	14	BQ805800	BQ805800
	20	179	37.5	360	9	AJ479969	AJ479969
	21	179	37.5	420	9	AJ479968	AJ479968
	22	178.2	37.4	480	9	AJ479970	AJ479970
C	23	178.2	36.2	240	9	AJ481887	AJ481887
C	24	172.8	36.2	240	9	AJ484886	AJ484886
C	25	172.8	36.2	556	9	AJ435524	AJ435524
C	26	172	36.1	240	9	AJ484593	AJ484593
	27	168	35.2	360	9	AJ485127	AJ485127
	28	166.4	34.9	480	9	AJ479971	AJ479971
	29	162.4	34.0	516	14	BQ755868	BQ755868
	30	161.4	33.8	405	14	BQ755493	BQ755493
	31	161.4	33.8	480	9	AJ481886	AJ481886
	32	159.8	33.5	364	10	BE423915	BE423915
	33	158	33.1	742	14	BQ249043	BQ249043
	34	150.8	31.6	420	9	AJ485287	AJ485287
	35	150	31.4	606	14	BQ245034	BQ245034
	36	150	31.4	818	14	BQ806739	BQ806739
	37	148.4	31.1	497	12	BF293550	BF293550
	38	148.4	31.1	515	14	BQ606957	BQ606957
	39	148.4	31.1	518	10	BE402654	BE402654
	40	148.4	31.1	518	14	BQ608417	BQ608417
	41	148.4	31.1	574	14	BQ805851	BQ805851
	42	148.4	31.1	584	14	BQ244692	BQ244692
	43	148.4	31.1	592	14	BQ246427	BQ246427
	44	148.4	31.1	640	10	BE427016	BE427016
	45	148.4	31.1	640	10	BE427051	BE427051

ALIGNMENTS

RESULT 1
LOCUS BQ755360 567 bp mRNA linear EST 26-JUL-2002
DEFINITION EBbed07_SQ002.L09.R endospERM, 28 DPA, no treatment, cv Optic,
EBbed07_Hordeum vulgare cDNA clone EBbed07_SQ002.L09.5', mRNA

ACCESSION BQ755360
VERSION BQ755360.1 GI:21963832
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
; Trilicaceae: Hordeum.
1 (bases 1 to 567)

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT Contact: Maugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estscri.sari.ac.uk.

FEATURES location/Qualifiers
1..567
/organism="Hordeum vulgare"

```

/cultivar="Optic"
/db.xref="taxon:4513"
/clone="EBed07.S0002.L09"
/clone_lib="endosperm, 28 DPA, no treatment, cv Optic, Ebed07"
/tissue_type="endosperm"
/dev_stage="28 DPA"
/lab_host="DH10B"
/Note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from endosperm tissue dissected from developing grains (28 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."
Gene Function) project."

BASE COUNT      197 a      224 c      48 g      98 t
ORIGIN
Query Match      67.4%; Score 321.6; DB 14; Length 567;
Best Local Similarity 84.8%; Pred. No. 1e-59;
Matches 373; Conservative 0; Mismatches 64; Indels 3; Gaps 1;

QY 4 AGGCACTAAACCTTGACGCAAGTTGCATTCACACACATCATATCTGCCGAG 63
    |||||||
D6 AGGCACTAAACCTTGACGCAAGTTGCATTCACACACATCATATCTGCCGAG 165
    |||||||
QY 64 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGCAGCAACCGTTTAC 123
    |||||||
D6 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGCAGCAACCGTTTAC 225
    |||||||
QY 124 ACACCCCAACAATTTTCCCTATCTACGAGGAATTTGCCCAATTCAAATACCA 183
    |||||||
D6 ACACCCCAACAATTTTCCCTATCTACGAGGAATTTGCCCAATTCAAATACCA 285
    |||||||
QY 184 ACCCCCTACACCAACAACATTCGCCCAACACACACACACACCTTCTCGGCC 243
    |||||||
D6 ACCCCCTACACCAACAACATTCGCCCAACACACACACACACCTTCTCGGCC 345
    |||||||
QY 244 CAACACACATTTCCCTGCAACCAACACATTTCCCGCCCAAGAACATTTGCC 303
    |||||||
D6 CAACACACATTTCCCTGCAACCAACACATTTCCCGCCCAAGAACATTTGCC 405
    |||||||
QY 304 CAACACATTTCCCTGCAACCAACACATTTCCCGCCCAAGAACATTTCCA 363
    |||||||
D6 CAACACATTTCCCTGCAACCAACACATTTCCCGCCCAAGAACATTTCCA 465
    |||||||
QY 364 CAATATATTTTTCAGCAACCAACCAATCATACCTGTGCAACCTCAACGCCATTTCT 423
    |||||||
D6 CAATATATTTTTCAGCAACCAACCAATCATACCTGTGCAACCTCAACGCCATTTCT 522
    |||||||
QY 424 CAACACCTCAACCAAGTCCC 443
    |||||||
D6 CAACACCTCAACCAAGTCCC 542
    |||||||

RESULT 2
LOCUS BE195280 585 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMH0088M16f Hordeum vulgare 5-45 DAP spike EST library
            HVCDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0088M16f,
            mRNA sequence.
ACCESSION BE195280
VERSION BE195280.3 GI:16321166
KEYWORDS EST
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
REFERENCE 1 (bases 1 to 585)
AUTHORS Wing,R., Close,T.J., Kleinjofs,A., Wise,R., Begum,D., Fritsch,D., Yu
        ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
        ,R.D., Close,S.J., Oates,R. and Main,D.

```

```

TITLE
JOURNAL Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 5-45 DAP spike cDNA library
COMMENT Unpublished (2001)
            On Jun 26, 2000 this sequence version replaced gi:13188066.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hg bases = 341
            Seq primer: AATTAACCTCCTAAAGG
            High quality sequence stop: 574.
            Location/Qualifiers
            1..585
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db.xref="taxon:4513"
            /clone="HVSMH0088M16f"
            /clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
            HVCDNA0009 (5 to 45 DAP)
            /tissue_type="5-45 DAP Spike"
            /lab_host="SOLR"
            /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmid SK(-) cDNA phagemids (Choi) in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinjofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/gqpages/bgn/31/cover.html)"

BASE COUNT      208 a      228 c      47 g      102 t
ORIGIN
Query Match      66.9%; Score 319; DB 10; Length 585;
Best Local Similarity 84.8%; Pred. No. 3.7e-59;
Matches 363; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY 4 AGGCACTAAACCTTGACGCAAGTTGCATTCACACACATCATATCTGCCGAG 63
    |||||||
D6 AGGCACTAAACCTTGACGCAAGTTGCATTCACACACATCATATCTGCCGAG 177
    |||||||
QY 64 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGCAGCAACCGTTTAC 123
    |||||||
D6 CCATATCCACAAAACCATATCTACCGCAAAACCATTTCCAGTGCAGCAACCGTTTAC 237
    |||||||
QY 124 ACACCCCAACAATTTTCCCTATCTACGAGGAATTTGCCCAATATCAATACCA 183
    |||||||
D6 ACACCCCAACAATTTTCCCTATCTACGAGGAATTTGCCCAATATCAATACCA 297
    |||||||
QY 184 ACCCCCTTACACCAACACATTTCCCGCCCAACCAACCAACCTTCTCGGCC 243
    |||||||
D6 ACCCCCTTACACCAACACATTTCCCGCCCAACCAACCAACCTTCTCGGCC 357
    |||||||
QY 244 CAACACATTTCCCTGCAACCAACCAATTTCCCGCCCAACCAACCAATTTCC 303
    |||||||

```


Fax: 864 656 4293
Email: twing@clermont.edu
Total bp bases = 304
Seq primer: AATTAACTTCACTAAAGCG
High quality sequence start: 8
High quality sequence stop: 828.

FEATURES

Source

Location/Qualifiers

1..936

/organism="Hordeum vulgare"

/cultivar="Morex"

/db_xref="taxon:4513"

/clone="HYSMEL0004123f"

/clone_11b="Hordeum vulgare 20 DAP spike EST library"

/tissue_type="20 DAP spike"

/lab_host="SOLR"

/note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, St Close, Ty Close). Whole spikes with awns trimmed were collected at 20 DAP (Fenton). Total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give Bluescript SK(-) cDNA phagemids in the Ty Close lab at the University of California, Riverside (Choi). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wang). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wang, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders. Also see Close Ty, Wang R, Kleinholz A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/g99pages/g99/31/cover.html)"

BASE COUNT

338 a 351 c 69 g 177 t 1 others

Query Match

Best Local Similarity 81.4%; Score 287.4; DB 12; Length 936;

Matches

346; Conservative 0; Mismatches 76; Indels 3; Gaps 1;

QY 30 GTTGCAATCAACCAATATATCTGCCGAGCATATCCAAAACCCATATAC 89
DB 3 GTTGCAATCAACCAATATATCTGCCGAGCATATCCAAAACCCATATAC 62
QY 90 GCAAAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149
DB 63 GCAAAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 122
QY 150 ACCAGAGAAATGTTTCCCAATATCAAAACCCCTTACACCAACCAACCAAT 209
DB 123 ACCAGAGAAATGTTTCCCAATATCAAAACCCCTTACACCAACCAACCAAT 182
QY 210 CCCCCAACCAACCAACCAATCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 269
DB 183 CCCCCAACCAACCAACCAATCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 242
QY 270 ACAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 329
DB 243 ACAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 302
QY 330 ACAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 389
DB 303 ACAACCAATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 362
QY 390 ATCATACCTGTCGAACCTCAAGCCATTTCTTCAACCAACCAACCAACCA 449
DB 363 ACCCATGCC--ACAACCAACCAACCAATACCAACCAACCAACCAACCAACCA 419

QY 450 AGCTT 454
DB 420 ACCCT 424

RESULT 5

LOCUS

AJ435523/c

DEFINITION

AJ435523 S00011 Hordeum vulgare cDNA clone S000110053F10F1, mRNA

ACCESSION

AJ435523

VERSION

AJ435523.1

KEYWORDS

EST

SOURCE

Hordeum vulgare.

ORGANISM

Hordeum vulgare.

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

AUTHORS

1 (bases 1 to 515)

TITLE

Barley EST's

JOURNAL

Unpublished (2002)

COMMENT

Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O.Box 56 (Vilkkinkari 6A), University of Helsinki FIN-00014, Finland.

FEATURES

Location/Qualifiers

Source

1..515

organism

"Hordeum vulgare"

db_xref

"taxon:4513"

clone

"S0001100053F10F1"

dev_stage

"S00011"

note

"Developing seed"

BASE COUNT

94 a 66 c 184 g 171 t

Query Match

Best Local Similarity 85.3%; Score 246.8; DB 9; Length 515;

Matches

297; Conservative 0; Mismatches 27; Indels 24; Gaps 1;

QY 4 AGCACTAAACCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 63
DB 345 AGCACTAAACCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 286
QY 64 CCATATCAACCAACCAATATCTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 123
DB 285 CCATATCAACCAACCAATATCTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 226
QY 124 ACACCCCAACCAATATCTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 183
DB 225 ACACCCCAACCAATATCTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 166
QY 184 ACCCCCTTACCAACCAACCAATCTCCCAACCAACCAACCAACCAACCAACCA 243
DB 165 ACCCCCTTACCAACCAACCAATCTCCCAACCAACCAACCAACCAACCAACCA 130
QY 244 CAACCAACCAATCTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 303
DB 129 CAACCAACCAATCTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 22
QY 304 CAACCAACCAATCTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 351
DB 69 CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 22
RESULT 6
LOCUS
DEFINITION
A1485191 420 bp mRNA linear EST 24-MAY-2002
A1485191 S00011 Hordeum vulgare cDNA clone S000110016F01F1, mRNA
sequence.

ACCESSION AJ485191 GI:21201146
 VERSION AJ485191.1
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 420)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's
 JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..420
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100164F01F1"
 /clone_lib="S00011"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

BASE COUNT 145 a 156 c 43 g 76 t

ORIGIN

Query Match 51.6%; Score 246.2; DB 9; Length 420;
 Best Local Similarity 87.4%; Pred. No. 1.9e-43;
 Matches 291; Conservative 0; Mismatches 18; Indels 24; Gaps 1;

QY 4 AGGCACTAAACCCCTTGAGCAGCAAGTGTGATCAGCAACATCATATCTGCGCAG 63
 DB 111 AGGCACTAAACCCCTTGAGCAGCAAGTGTGATCAGCAACATCATATCTGAGCAG 170
 DB 171 CCATATCCCAAAACCCATATCTACCGCAACATATTTCCAGTGACCAACCGTTTCAC 230
 QY 64 CCATATCCCAAAACCCATATCTACCGCAAAACATTTCCAGTGACCAACCGTTTCAC 123
 DB 171 CCATATCCCAAAACCCATATCTACCGCAACATATTTCCAGTGACCAACCGTTTCAC 230
 QY 124 ACACCCCAACATATTTCCCTATCTACGAGAGATTTTCCCATATCAATATGCA 183
 DB 231 ACACCCCAACATATTTCCCTATCTACGAGAGATTTTCCCATATCAATATGCA 290
 QY 184 ACCCCCTTACAGCAGCAACATTTCCCGCAACAGCAACATCTTTCTCGGCGC 243
 DB 291 ACCCCCTT-----ACAACGCAACATCTTTCTCGGCGC 326
 QY 244 CAACACCATTTCCCTGCGCAACGCAACATTTCCCGACGCGCAAGCAATTCGC 303
 DB 327 CAACACCATTTCCCTGCGCAACGCAACATTTCCCGACGCGCAAGCAATTCGC 386
 QY 304 CAACACCATTTCCCTGCGCAACGCAACGCAACGCA 336
 DB 387 CAGCAACGCAACGCAACATTTCCGACGCAACGA 419

RESULT 7
 LOCUS AJ484592 360 bp mRNA linear EST 24-MAY-2002
 DEFINITION AJ484592 S00011 Hordeum vulgare cDNA clone S0001100047B07F1, mRNA sequence.
 ACCESSION AJ484592
 VERSION AJ484592.1 GI:21200548
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 360)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's

JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..360
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100047B07F1"
 /clone_lib="S00011"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

BASE COUNT 67 a 37 c 136 g 120 t

ORIGIN

Query Match 51.5%; Score 245.8; DB 9; Length 360;
 Best Local Similarity 87.6%; Pred. No. 2.2e-43;
 Matches 290; Conservative 0; Mismatches 17; Indels 24; Gaps 1;

QY 4 AGGCACTAAACCCCTTGAGCAGCAAGTGTGATCAGCAACATCATATCTGCGCAG 63
 DB 309 AGGCACTAAACCCCTTGAGCAGCAAGTGTGATCAGCAACATCATATCTGAGCAG 250
 QY 64 CCATATCCCAAAACCCATATCTACCGCAAAACATTTCCAGTGACCAACCGTTTCAC 123
 DB 249 CCATATCCCAAAACCCATATCTACCGCAACATATTTCCAGTGACCAACCGTTTCAC 190
 QY 124 ACACCCCAACATATTTCCCTATCTACGAGAGATTTTCCCATATCAATATGCA 183
 DB 189 ACACCCCAACATATTTCCCTATCTACGAGAGATTTTCCCATATCAATATGCA 130
 QY 184 ACCCCCTTACAGCAGCAACATTTCCCGCAACGCAACATCTTTCTCGGCGC 243
 DB 129 ACCCCCTT-----ACAACGCAACATCTTTCTCGGCGC 94
 QY 244 CAACACCATTTCCCTGCGCAACGCAACATTTCCCGACGCGCAAGCAATTCGC 303
 DB 93 CAACACCATTTCCCTGCGCAACGCAACATTTCCCGACGCGCAAGCAATTCGC 34
 QY 304 CAACACCATTTCCCTGCGCAACGCAACGCAACGCA 334
 DB 33 CAACACCAACGCAACGCAACGCAACGCAACGCA 3

RESULT 8
 LOCUS AJ484591 300 bp mRNA linear EST 24-MAY-2002
 DEFINITION AJ484591 S00011 Hordeum vulgare cDNA clone S0001100043B09F1, mRNA sequence.
 ACCESSION AJ484591
 VERSION AJ484591.1 GI:21200547
 KEYWORDS EST.
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 300)
 Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
 TITLE Barley EST's
 JOURNAL Unpublished (2002)
 COMMENT Contact: Schulman AH
 Institute of Biotechnology
 University of Helsinki
 P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014, Finland.

FEATURES
 source Location/Qualifiers
 1..300
 /organism="Hordeum vulgare"
 /db_xref="taxon:4513"
 /clone="S0001100043B09F1"

BASE COUNT 49 a 26 c 116 g 109 t
 ORIGIN /clone.lib="S00011"
 /dev_stage="Developing seed"
 /note="12,15,18 days after pollination"

Query Match 49.6%; Score 236.8; DB 9; Length 300;
 Best Local Similarity 87.3%; Pred. No. 1.8e-41;
 Matches 281; Conservative 0; Mismatches 17; Indels 24; Gaps 1;

QY 13 AACCTTGACGCAAGAGTTGCAATACCAACAATATATCGCCGACGCAATATCA 72
 Db 300 AACCTTGACGCAAGAGTTGCAATACCAACAATATATCGAGGACGCAATATCA 241
 QY 73 CAAACCCATATACGCAAGAGTTGCAATACCAACAATATATCGCCGACGCAATATCA 132
 Db 240 CAAACCCATATACGCAAGAGTTGCAATACCAACAATATATCGAGGACGCAATATCA 181
 QY 133 CAATATTTTCCCTATATACGAGAGATTTTCCCATATATCAATATCAACCCCTA 192
 Db 180 CAATATTTTCCCTATATACGAGAGATTTTCCCATATATCAATATCAACCCCTA 122
 QY 193 CAACCAACAACCAATTTCCCAACAACAACCACTCTCTCGGCGCCCAACAACA 252
 Db 121 -----ACAACCAACAACCACTCTCTCGGCGCCCAACAACA 85
 QY 253 TTCCCTTGACGCAAGAGTTTCCGACGCGCCCAACAACAATTTCCCAACAACA 312
 Db 84 TTCCCTTGACGCAAGAGTTTCCGACGCGCCCAACAACAATTTCCCAACAACA 25
 QY 313 TTCCCTTGACGCAAGAGTTTCCGACGCGCCCAACAACAATTTCCCAACAACA 334
 Db 24 CAACCAACAACGCAACAACAAC 3

RESULT 9
 LOCUS BQ609718 940 bp mRNA linear EST 25-JUN-2002
 DEFINITION BRY.5804 wheat EST endosperm library Triticum aestivum cDNA 5',
 mRNA sequence.

ACCESSION BQ609718
 VERSION BQ609718.1 GI:21559057
 KEYWORDS EST
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 TITLE 1 (bases 1 to 940)
 AUTHORS Clarke,B., Lambrecht,M. and Rhee,S.
 JOURNAL Assessing the utility of Arabidopsis genomic information for
 COMMENT Interpreting wheat EST sequences
 Unpublished (2002)
 Contact: Lambrecht M
 The Arabidopsis Information Resource
 Carnegie Institution of Washington, Dept. of Plant Biology
 260 Panama Street, Stanford, CA 94305, USA
 Tel: 1 650 325 1521 x 251
 Fax: 1 650 325 3748
 Email: rhee@croma.stanford.edu.

FEATURES
 source Location/Qualifiers
 1..940
 /organism="Triticum aestivum"
 /cultivar="WYuna"
 /db_xref="taxon:4565"
 /clone_lib="wheat EST endosperm library"
 /tissue_type="endosperm"
 /dev_stage="developing endosperm tissue 8, 10 and 12 DPA
 (days post anthesis)"
 /note="Vector: Bluescript II SK(-)"

BASE COUNT 340 a 365 c 90 g 145 t
 ORIGIN

Query Match 47.0%; Score 224.2; DB 14; Length 940;
 Best Local Similarity 73.5%; Pred. No. 1.3e-38;
 Matches 324; Conservative 0; Mismatches 93; Indels 24; Gaps 2;

QY 4 AGCAACTTAACCCCTGAGCAGAGTTGCAATACCAACAACAATATATCGCCGAG 63
 Db 166 AGCAACTTAACCCCTGAGCAGAGTTGCAATACCAACAACAATATATCGCCGAG 210
 QY 64 CCATATCCCAAAACCCATATATACGCAAGAGATTTTCCCATATATCAATATCA 123
 Db 211 CCGTATCCCAAAACCCATATATACGCAAGAGATTTTCCCATATATCAATATCA 270
 QY 124 ACACCCCAACAATATTTCCCTATATACGAGAGATTTTCCCATATATCAATATCA 183
 Db 271 ACACCCCAACAATATTTCCCTATATACGAGAGATTTTCCCATATATCAATATCA 330
 QY 184 ACCCCCTTACACCAACAACAACATTTCCCAACAACAACATTTCCCTTCCGCCC 243
 Db 331 ACCCCCTTACACCAACAACAACATTTCCCAACAACAACATTTCCCTTCCGCCC 390
 QY 244 CAACAACCAATTTCCCTGCAACAACAACAATTTCCCAACAACAATTTCCCTG 303
 Db 391 CAACAACCAATTTCCCTGCAACAACAACAATTTCCCAACAACAATTTCCCTG 450
 QY 304 CAACAACA-----TTCCCTGCAACAACAACAATTTCCCAACAACA 354
 Db 451 CAACAACAACAATTTCCCTGCAACAACAACAATTTCCCAACAACAATTTCCCTG 510
 QY 355 CCATTTCAACAATATTTTCCCAACAACAACAATTTCCCAACAACAATTTCCCTG 414
 Db 511 CAACAACAACAATATTTTCCCAACAACAACAATTTCCCAACAACAATTTCCCTG 570
 QY 415 CCATTTCCCAACAACAATTTTCCCAACAACAACAATTTCCCAACAACAATTTCCCTG 435
 Db 571 CCATTTCCCAACAACAATTTTCCCAACAACAACAATTTCCCAACAACAATTTCCCTG 591

RESULT 10
 LOCUS BQ472381 595 bp mRNA linear EST 31-MAY-2002
 DEFINITION HB09M16T BC Hordeum vulgare cDNA clone HB09M16 5-PRIME, mRNA
 sequence.

ACCESSION BQ472381
 VERSION BQ472381.1 GI:21284510
 KEYWORDS EST
 SOURCE Hordeum vulgare.
 ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 TITLE 1 (bases 1 to 595)
 AUTHORS Radchuk,V., Zhang,H., Weschke,M., Potokina,E. and Wobus,U.
 JOURNAL Barley ESTs from developing seeds
 COMMENT Unpublished (2002)
 Contact: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3 06466, Gatersleben, Germany
 Tel: 039462-5522
 Fax: 039462-5595
 Email: stein@ipk-gatersleben.de
 Insert length: 595 Std Error: 0.00
 Plate: 9 row: M column: 16
 Seq primer: 73.

FEATURES
 source Location/Qualifiers
 1..595
 /organism="Hordeum vulgare"
 /cultivar="Baiké"
 /db_xref="taxon:4513"
 /clone_lib="HB09M16"
 /tissue_type="developing caryopsis"
 /dev_stage="8-15 DAP (days after pollination)"

BASE COUNT 340 a 365 c 90 g 145 t
 ORIGIN


```

Db      409 CATTCCCTAGACACCAACCAAAATATTTCCCAACACCCCAACCAACCATTCCTCCCTG 468
      407 CTCACACAGCATTT 420
      526 CACACACACCATTT 539

RESULT 12
B0245760 560 bp mRNA linear EST 03-MAY-2002
LOCUS     TAE15020E06R TAE15 Triticum aestivum cDNA clone TAE15020E06R, mRNA
DEFINITION
ACCESSION B0245760
VERSION    B0245760.1 GI:20441636
KEYWORDS   EST
SOURCE     bread wheat.
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Triticum.
            1 (bases 1 to 560)
REFERENCE 1 Cloutier S.
            Wheat functional genomics - glenlea developing seeds cDNA libraries
            Unpublished (2002)
            Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dufferin Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutiere@agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >1.4 kb
            Plate: 020 row: E column: 06
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..560
                /organism="Triticum aestivum"
                /cultivar="Glenlea"
                /db_xref="taxon:4565"
                /clone="TAE15020E06R"
                /clone_1ib="TAE15"
                /tissue_type="developing seeds"
                /dev_stage="15 days after anthesis"
                /lab_host="E. coli DH10B"
                /note="Vector: PCMV-SPORT6.0 (Invitrogen Technologies);
                Site_1: Nott; Site_2: MUI; mRNA obtained from wheat seeds
                of cultivar Glenlea 15 days post-anthesis"
BASE COUNT 192 a 240 c 41 g 87 t
ORIGIN
Query Match 45.7%; Score 218; DB 14; Length 560;
Best Local Similarity 74.1%; Pred. No. 2,5e-37;
Matches 329; Conservative 0; Mismatches 85; Indels 30; Gaps 3;

OY      25 CAAGAGTTCATACACCAACATATCTCCGACGATATTCACAAACCAATAT 84
      68 CATATCTTTTCCATACACCAACATATTCACAGAGCATATTCACAAACCAATAT 127
OY      85 CTAACCAACCAATTCAGTACAGACGATTCACAGCCCAACCAATATTTCC 144
      128 TCATCAGACACCAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 187
OY      145 TATTCACAGAGAAATGTTCCCAATATTCATTCATTCATTCATTCATTCATTCATTC 204
      188 GAGCATTCACCAACCAATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 247
OY      205 CCAATTC-----CCCAACCAACCAACCAACCAACCAACCAACCAACCAATTC 255

```

```

Db      248 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAATTC 307
      256 CCGTGGACAGCAACCAACCAATTCCTCCAGCCCAAGAACCAAT-----TCCCA 306
      308 CCGTGGACAGCAACCAACCAATTCCTCCAGCCCAAGAACCAATTCCTCCAGCAACCA 367
OY      307 CAACCAATTCCTCCAGCAACCAACCAATTCCTCCAGCCCAAGAACCAATTC----- 361
      368 CAGCATTCCTCCAGCAACCAACCAATTCCTCCAGCCCAAGAACCAATTCCTCCAGCAAC 427
OY      362 -----ACCAATATTCCTCCAGCAACCAACCAATTCCTCCAGCCCAAGAACCAAG 414
      428 CATTCACCAACCAATATTCCTCCAGCAACCAATTCCTCCAGCAACCAATTCCTCCAGCAAC 487
OY      415 CCAATTCCTCCAGCAACCAACCAATTCCTCCAGCAACCAATTCCTCCAGCAACCAACCA 438
      488 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 511
      488 CCAATTCCTCCAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 511

```

```

RESULT 13
B0472392 611 bp mRNA linear EST 31-MAY-2002
LOCUS     HB09M02T BC Hordeum vulgare cDNA clone HB09M02.5-PRIME, mRNA
DEFINITION
ACCESSION B0472392
VERSION    B0472392.1 GI:21284521
KEYWORDS   EST
SOURCE     Hordeum vulgare.
ORGANISM   Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Hordeum.
            1 (bases 1 to 611)
REFERENCE 1 Radchuk V., Zhang H., Wesche W., Potokina E. and Wobus U.
            Barley ESTs from developing seeds
            Unpublished (2002)
            Contact: Steffen Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: steinle@ipk-gatersleben.de
            Insert Length: 611 Std Error: 0.00
            Plate: 9 row: M column: 2
            Seq primer: 73.
            Location/Qualifiers
                1..611
                /organism="Hordeum vulgare"
                /cultivar="Barke"
                /db_xref="taxon:4513"
                /clone="HB09M02"
                /clone_1ib="BC"
                /tissue_type="developing caryopsis"
                /dev_stage="8-15 DAP (days after pollination)"
                /lab_host="X110-Gold"
                /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                cDNA); Site_2: XhoI (3'-end of cDNA); developing caryopsis
                , 8-15 DAP(days after pollination) Due to a cloning
                artefact caused by the kit, in most cases the EcoRI site
                is NOT present, as well as the EcoRI site used for
                cloning. To excise the insert, restriction sites upstream
                EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
                due to the cloning system used blue/white selection for
                recombinants is not 100% reliable."
BASE COUNT 198 a 244 c 57 g 112 t
ORIGIN
Query Match 45.7%; Score 217.8; DB 14; Length 611;
Best Local Similarity 71.3%; Pred. No. 2,9e-37;
Matches 323; Conservative 0; Mismatches 112; Indels 18; Gaps 2;

```

OY	11	TTAAACCTTGGACGCCAGATGTTGGCAATTCACACCAACATATATATCGCGCAGACCATTC	70
Db	155	TAAACCGTATGCTACCAAGAGTTGGCATTCACACCAACACCATTTTGTGAACCAACATCTAT	214
OY	71	CACAAAACCCATATCTACCGCAAAAACCATTTCCATGTCAGCAACCGTTTCACACACCC	130
Db	215	ATCTGAGCAACCATATCTACACCAACCATATCTACCGCAGCAACCATTTCCGACCC	274
OY	131	AACAATATTTCCGCTATCTACCAAGAGATTTGTTCCGCAATATGCAATAATCAACACCCCC	190
Db	275	AACAATTTTCCCTATCTACCAACGCAACCATTTCCGCAATCCCAACAACACCCCC	334
OY	191	TACACACCAACACCATTC-----CCCAACAACCAACAACCTCTTCTCGCG	241
Db	335	TACAACCAACAACCATTTCCCGCTGCAACCGCAACCAACAACATCTTCTCGAGT	394
OY	242	CCCAACAACCATTTCCCGCTGCAACCAACAACCATTTCCCGACCCCAAGAACCATTC	301
Db	395	CCCAACAACCATTTCACTGGCAACCAACAACCTTTTCCGACCCCAACAACCATTC	454
OY	302	CC-----CAACAACCATTTCCCGCTGCAACCAACAACCATTTCCCGACCCCAAG	352
Db	455	CCGAGCAACCAACAACCGCTTCCCGCTAGCAACCAACAACAATATTTCCCAACAACAC	514
OY	353	AACCAATTCACAACAATATATTTTCAGCAACCCCAACAATCATATACCTGTGCAACTGAC	412
Db	515	AACAACCATTTCTCCGCAACCCCAACAACCAACATCATCAACCTTTCTCGGCGCCCAACAAC	574
OY	413	AGCATTTCCTCAACAACCTCAACAGAGCGCCG	445
Db	575	CATTCCCTGGCCACCAACAACCATTTTTCG	607

RESULT 14
 LOCUS B1951055
 DEFINITION
 854 bp mRNA linear EST 19-OCT-2001
 HVSME10024C19f Hordeum vulgare spike EST library HVCNDA0012
 (Fusarium infected) Hordeum vulgare cDNA clone HVSME10024C19f, MRNA
 sequence.
 B1951055.
 B1951055.1 GI:16293688
 EST.
 Hordeum vulgare.
 Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 1 (bases 1 to 854)
 Wang,R., Mehnbauer,G.J., Close,T.J., Kleinhoft,A., Wise,R., Heinen
 S., Begum,D., Fritsch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
 Simmons,J., Fenton,R.D., Malvarresi,M., Choi,D.W., Oates,R. and Main
 D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Fusarium Infected Morex spike cDNA library
 Unpublished (2001)
 Contact: Wang RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 total hg bases = 508
 Seq primer: ATTATACCTCCTCAATAGG
 High quality sequence start: 3
 High quality sequence stop: 656.
 Location/Qualifiers
 1..854

```

FEATURES
source
Location/Qualifiers
1..854
/organism="Hordium vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HvSMEL0024C19f"
/clone_lib="Hordium vulgare spike EST library HvCDNA0012

```

(Fusarium infected)"
/fissure_type="Spike"
/lab_host="Jrc121"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plants were grown at the university of Minnesota in the GJ Muehlbauer lab; spikes were harvested and snap frozen at 0, 1, 2, 3, 4, 5, 6, and 8 days after Fusarium graminearum inoculation (Heinen). In the TJ Close lab at the university of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all eight RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi, Fenton, Malatrasi). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinbotts A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

	Query Match	45.1%	Score 215	DB 13	Length 854
	Best Local Similarity	73.7%	Pred. 1.3e-36		
	Matches 353	Conservative	0	Mismatches 95	Indels 31
				Gaps	5
QY	4	AGGCACTAATACCCCTTCGAGCGAAGCTTGCATACACC-----ACAA	45		
Db	33	AGGCACTAATACCCCTTCGAGCGAAGCTTGCATACACC-----ACAA	92		
QY	46	CAATCATATCTGCGCGAGGCATATTCACAAACCATATTCACGCGAAACCATTTGCCA	105		
Db	93	CAATCATATCTGAGGAGGCATATTCATTAACCCATATTTACGCGAACCATTTGCCA	152		
QY	106	GTGACAGCAACCGTTTTCACACACCCCAACATATTTTCCCTATTCACAGAGAAATTTGTT	165		
Db	153	GTGACAGCAACCATTTTCCACACCCCGACACATATTTTCCCTATTCACAGAGAAATTTGTT	212		
QY	166	CCGCAATATCAATATACCAACCCCTCTACACACGACAGACATTTCCG--CAACACACAA	224		
Db	213	CCGCAATATCAATATACCAACCCCTCTACACACGACAGACATTTCCG--CAACACACAA	272		
QY	225	ACAACGCTTCTCCCTCGGCC---CCACACACATTTCCCTCTGCGACACCAACACATTTCC	281		
Db	273	ACCATTTCCCAACAGACCCCATTCGCCACACACGACACACATATTCACACACACAAAC	332		
QY	282	CCAGCCCAAGAACCAATTTCCCAACACCAATTTCCG---CTGGCAACACACACATTC	338		
Db	333	ATTTTCCGCAACACCCATTCGCCACACACCAACCAACCAATTCGCCACACACACCAATTC	392		
QY	339	TCCGACGCCCAAGAACCAATTCACACCAATATTTTCCAGACACACCCCA-----CAATC	392		
Db	393	TCCGACCAACACCCCTTCGATTCACACACCAATTTTCCCAACACACCAATTTTGGGACA	452		
QY	393	ATTACCTGTGCAACCTTCACACAGGCATTTTCTCAGACACGACGATGCCGACACAG	451		
Db	453	ACAACGATTTCTATTCGACAGCAACCAATTCACAGACAGACCAACGACACTCTCTACAAG	511		

RESULT 15

BOH04665

LOCUS

DEFINITION

BOH04665 706 bp mRNA linear EST 31-JUL-2002

WH3557_002.H03ZS wheat developing grains cDNA library Triticum

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:35:40 : Search time 48 Seconds

(without alignments)
438.617 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906
Sequence: 1 MRQINPCSELOSPOQSYLQ.....PPQQPQPPVPOQASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	906	100.0	158	21	AAV54568 A synthetically tr
2	262.5	29.0	1162	21	AAV96525 Kaposi's sarcoma-a
3	262.5	29.0	1162	21	AAV58500 HHV8 ORF 73 protei
4	262.5	29.0	1162	22	AAAB62351 Amino acid sequenc
5	262.5	29.0	1162	23	AAAB05621 Kaposi's sarcoma-a
6	255.5	28.2	369	19	AAAB62647 Mature durum wheat
7	244.5	27.0	1362	22	AAU38416 Salmorella typhi c
8	234	25.8	266	22	AAU01799 Wheat A-glialdin.
9	217	24.0	1329	22	AAU34519 E. coli cellular p
10	213	23.5	308	21	AAAB4459 Plant viral moveime

11	204	22.5	1299	22	AAAB24322 Human EST encoded
12	204	22.5	1404	13	AAAB26049 MSF precursor. Sy
13	204	22.5	1404	22	AAAB60568 Human megakaryocyt
14	204	22.5	1404	22	AAAB29773 Human megakaryocyt
15	204	22.5	1415	22	AAU32262 Novel human secret
16	202.5	22.4	717	22	ABG17438 Novel human diagno
17	200.5	22.1	647	22	ABAB7093 Drosophila melanog
18	200.5	22.1	672	22	ABAB59473 Drosophila melanog
19	196.5	21.7	2285	22	ABAB3057 Drosophila melanog
20	196	21.6	757	22	AAAG5017 Shrimp white spot
21	195	21.5	446	22	ABAB70663 Drosophila melanog
22	194	21.4	882	22	ABAB51046 Drosophila melanog
23	190	21.0	1798	22	ABAB71695 Drosophila melanog
24	187	20.6	526	21	ABAB12717 Human ORF2828
25	187	20.6	987	21	AAAB43064 Human tyrosine pho
26	187	20.6	1253	21	AAAB29663 Human tyrosine pho
27	187	20.6	1264	21	AAAB29664 Human tyrosine pho
28	187	20.6	1274	20	AAAB9253 Human Alp. Homo s
29	187	20.6	1583	21	AAAB29662 Human tyrosine pho
30	187	20.6	1636	21	AAAB29661 Human histidine do
31	184.5	20.4	450	22	ABAB71041 Drosophila melanog
32	184	20.3	667	19	AAAB48760 BOP1 protein. Mus
33	183.5	20.3	371	19	AAAB64535 Human leukocyte ce
34	183.5	20.3	371	23	ABAB61444 Human NF-kB activa
35	182.5	20.1	522	21	AAAG3630 Arabidopsis thalia
36	182.5	20.1	539	21	AAAG3629 Arabidopsis thalia
37	182.5	20.1	540	21	AAAG3628 Arabidopsis thalia
38	180	19.9	148	21	AAI92720 C. albicans hyphal
39	179.5	19.8	961	22	ABAB67326 Drosophila melanog
40	177.5	19.6	180	21	AAAB5073 Rat amelogenin pro
41	175	19.3	194	31	AAAB5072 Human polypeptide
42	175	19.3	748	22	AAAB42065 Human polypeptide
43	175	19.3	780	22	AAAB42065 Human polypeptide
44	174	19.2	439	13	AAAB28150 Sugar beet chitina
45	173.5	19.2	688	22	ABAB67404 Drosophila melanog

ALIGNMENTS

RESULT 1	AAV54568	AAV54568	standard. Protein: 158 AA.
ID	AAV54568	AAV54568	
AC	AAV54568	AAV54568	
XX	25-APR-2000	(first entry)	
DT	25-APR-2000	(first entry)	
XX	A synthetically truncated C hordein protein (glutenin).		
XX	C hordein gene: glutenin; seed storage protein: gluten; bread; pasta;		
KW	noode; breakfast cereal; snack food; cake; pastry; flour based sauce;		
KW	film; coating; adhesive; building material; packaging material; grain.		
XX	Hordeum vulgare.		
OS	Synthetic.		
XX			
FT	Key	Location/Qualifiers	
FT	Misc-difference 20	/note= "encoded by CCG"	
FT	Misc-difference 105	/note= "encoded by TTC"	
FT	Misc-difference 106	/note= "encoded by CCC"	
FT	Misc-difference 107	/note= "encoded by TGG"	
FT	Misc-difference 108	/note= "encoded by CAA"	
FT	Misc-difference 113	/note= "encoded by TTT"	
FT	Misc-difference 114	/note= "encoded by TTT"	
FT	Misc-difference 115	/note= "encoded by CCC"	

CC with the latency cycle of this virus. Potential antiviral treatments for
CC the above mentioned diseases may therefore be based on LANA deregulation.

XX Sequence 1162 AA:

Query Match 29.0%; Score 262.5; DB 21; Length 1162;

Best Local Similarity 46.8%; Pred. No. 5.9e-15;

Matches 74; Conservative 10; Mismatches 51; Indels 23; Gaps 8;

```

OY 2 ROLNPGSGE-LQSPQSYLQGPYQNPYLQKPPVQDPPFTPOQYFPLPEEL-FPOYQ 59
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 453 QQQQEPQDQEPQLQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 502
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 60 IFTPLQDQDPPFQQ--PQDPLPRPQDPPFQWQDQEPQDQEPQDQEPQDQEPQDQ 115
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 503 EPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 562
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 116 FPQD--PQDIIFFQDQDQSYPVQ--PQDPPQDQDPPVQ 149
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 563 EPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 597

```

RESULT 3

AAV58500
ID AAV58500 standard; Protein; 1162 AA.

XX AAV58500;

DT 10-APR-2000 (first entry)

XX HHV8 ORF 73 protein, SEQ ID NO:21.

DE HHV8 ORF 73 protein, SEQ ID NO:21.

XX HHV8; detection; diagnosis; Kaposi's sarcoma; AIDS; immunogen;

KW antigen.

XX Human herpesvirus type 8.

OS Human herpesvirus type 8.

FT Key Location/Qualifiers

FT Misc-difference 96 /Label= unknown

XX WO961909-A2.

XX 02-DEC-1999.

XX 26-MAY-1999; 99WO-US11407.

XX 26-MAY-1998; 98US-0086695.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Pau CP;

XX WPI: 2000-097142/08.

XX New methods and compositions for the detection of human herpesvirus

XX Claim 2; Page 59-62; 68pp; English.

XX Sequences AAV58480-Y58532 represent immunogenic polypeptides derived
CC from human herpes virus type 8 (HHV8, a gammaherpesvirus). HHV8 plays an
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The
CC invention relates to a novel method of detecting the presence of human
CC herpesvirus 8 in a biological sample using peptides representative of
CC dominant antigenic regions of HHV8. The method comprises contacting one
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing
CC biological sample, and detecting the formation of a complex between the
CC peptide and the antibody. The presence of a peptide-antibody complex
CC indicates the presence of human herpesvirus 8. The detection of HHV8
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The
CC HHV8-specific antibodies are useful therapeutically when for the passive
CC immunisation of a human against HHV8 infection, thereby reducing HHV8
CC related disease. The detection assays are highly specific, sensitive and

CC accurate. Early detection and treatment of Kaposi's sarcoma could
CC diminish the severity of symptoms related to AIDS and the sensitive
CC techniques could reduce erroneous characterisations of skin disorders.
CC Previous assays for HHV8 antibodies, such as immunofluorescence assays,
CC immunoblots and enzyme immunoassays lack the sensitivity and accuracy
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages
CC of the assays are that reproducible results are obtained and the method
CC is suitable for rapid throughput and screening of samples economically.

XX Sequence 1162 AA:

Query Match 29.0%; Score 262.5; DB 21; Length 1162;

Best Local Similarity 46.8%; Pred. No. 5.9e-15;

Matches 74; Conservative 10; Mismatches 51; Indels 23; Gaps 8;

```

OY 2 ROLNPGSGE-LQSPQSYLQGPYQNPYLQKPPVQDPPFTPOQYFPLPEEL-FPOYQ 59
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 453 QQQQEPQDQEPQLQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 502
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 60 IFTPLQDQDPPFQQ--PQDPLPRPQDPPFQWQDQEPQDQEPQDQEPQDQEPQDQ 115
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 503 EPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 562
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 116 FPQD--PQDIIFFQDQDQSYPVQ--PQDPPQDQDPPVQ 149
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 563 EPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQEPQDQ 597

```

RESULT 4

AAV62331
ID AAV62331 standard; Protein; 1162 AA.

XX AAV62331;

DT 29-JUN-2001 (first entry)

XX Amino acid sequence of KSHV tethering protein LANA.

DE Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;

KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;

KW KSHV; latency-associated nuclear antigen; LANA.

XX Kaposi's sarcoma associated herpesvirus.

OS WO200125484-A2.

XX 12-APR-2001.

XX 29-SEP-2000; 2000WO-US26908.

XX 01-OCT-1999; 99US-0410399.

XX (UNMI) UNIV MICHIGAN.

XX Robertson ES, Colter MA;

XX WPI: 2001-281736/29.

XX N-PSDB: AAF62901.

XX Disclosure; Fig 9B; 60pp; English.

XX The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operatively encoding a protein suitable
CC for tethering protein is LANA. The composition is useful in aiding the
CC tethering of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for

Db 39 000PCS0000QPPPLS0000QPPFS0000QPVLP00PFS000LPPFS0000QPPF 98
 QY 41 HPPQOYFPLPEE-LFPYOIPTPL0PQ0PPOQ-----PQ0PLPR0Q0PPM0PQ-QPP 94
 Db 99 S0000--PVLPPQ0PFS000LPPFS000PVLPPQ0PFS000QPPPS00LPPFS 156
 QY 95 Q0PEP-LP0P00PFP0P0P0P0Q011F0Q0P0QSYV0P0Q0PP-----Q0P0PV 148
 Db 157 0000PVLPP0P--PFS0000QPPF--PQ0PFS000PVL0001PFPVPSL00LNPCK 212
 QY 149 ---0QASCIWSM 157
 Db 213 VFLQ0CCSPWAM 224
 RESULT 7
 AAU38416
 ID AAU38416 standard: Protein: 1362 AA.
 AC AAU38416;
 DT 14-FEB-2002 (first entry)
 DE Salmonella typhi cellular proliferation protein #307.
 KW Antisense: prokaryotic cellular proliferation protein;
 KW antibiotic: antibacterial; drug design.
 OS Salmonella typhi.
 PN W0200170955-A2.
 PD 27-SEP-2001.
 PE 21-MAR-2001: 2001WO-US09180.
 PR 21-MAR-2000: 2000US-191078P.
 PR 23-MAY-2000: 2000US-206848P.
 PR 26-MAY-2000: 2000US-207727P.
 PR 23-OCT-2000: 2000US-242578P.
 PR 27-NOV-2000: 2000US-253625P.
 PR 22-DEC-2000: 2000US-257931P.
 PR 16-FEB-2001: 2001US-269308P.
 PA (ELIT-) ELITRA PHARM INC.
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
 PI Yamamoto RT, Xu HH;
 DR WPI: 2001-611495/70.
 DR N-PSDB: AAS56275.
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 14009; 511pp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 SO Sequence 1362 AA:
 Query Match 27.0%; Score 244.5; DB 22; Length 1362;
 Best Local Similarity 41.1%; Pred. No. 2,6e-13;
 Matches 74; Conservative 10; Mismatches 57; Indels 39; Gaps 11;
 QY 2 R0LNPSC0EL0SP0QSYLQ0PY-----P0NP-----YLP0KFPFV0QP 39
 Db 686 R0FAASQ0QRYSGE0PAC0P0PSLDDLF0SPMKVLYVEGH0EPLFTGVMPEST-PVQ0P 744
 QY 40 F-HTPQOYFPLPEELFPOYOIPTPL0PQ0PFP0PQ0PRL-PRPQ0PFPWQ0PQF-PQP 96
 Db 745 VAPQ0YQ0PVAPQ--PQY00PQ0PVAS0PQY00PQ0PVAPQ0Y0--0PQ0PVAPQ 798
 QY 97 Q0EPLPQ0PQ0PFP--Q0PQ0PFP0PQ011P0P0Q0SYV0PQ0PFP0PQ0PQV0QASCI 154
 Db 799 QYQ0PQ0PVAPQ0Y00PQ0PVAPQ0Q--YQ0PQ0--PVAPQ0Y00PQ0PAPQDLSLI 853
 RESULT 8
 AAU01799
 ID AAU01799 standard: Protein: 266 AA.
 AC AAU01799;
 DT 07-SEP-2001 (first entry)
 DE Wheat A-gliadin.
 KW Wheat: A-gliadin: epitope; coeliac disease; gluten intolerance;
 KW T-cell binding; antagonist; transglutaminase; transgenic plant.
 OS Triticum aestivum.
 PN W0200125793-A2.
 PD 12-APR-2001.
 PE 02-OCT-2000: 2000WO-GB03760.
 PR 01-OCT-1999: 99GB-0023306.
 PA (ISIS-) ISIS INNOVARTON LTD.
 PI Anderson RP, Hill AVS, Jewell DP;
 DR WPI: 2001-300179/31.
 PT Diagnosing coeliac disease or susceptibility to the disease in an
 PT individual, by detecting in vitro or in vivo T cells which bind
 PT immunodominant T cell epitope obtained from naturally occurring homolog
 PT of gliadin -
 PS Claim 1; Page 52; 107pp; English.
 CC The sequence represents wheat A-gliadin, A-gliadin derived peptides of
 CC the invention are used to test mammalian (preferably human)
 CC susceptibility to coeliac disease (gluten intolerance). The peptides are
 CC contacted with a blood sample and T cell recognition measured, a positive
 CC T-cell recognition indicating a susceptibility to coeliac disease. The
 CC peptides are useful for inducing tolerance in an individual and
 CC antagonists to the peptides are useful for treating or preventing coeliac
 CC disease in an individual and for producing an antibody specific to them
 CC or a wild-type sequence. A mutant gliadin protein (or its fragment of 15
 CC amino acids in length) whose wild-type sequence can be modified by
 CC transglutaminase to a sequence that comprises the epitope, but which has
 CC been modified in such a way that it does not contain sequence which can
 CC be modified by transglutaminase to a sequence that comprise the epitope
 CC is useful for decreasing the ability of gliadin protein to cause Coeliac

CC disease. Nucleic acids encoding proteins antagonistic to the T-cell
 CC binding of the epitopes are useful for obtaining a transgenic plant cell
 CC or seed and for the production of a protein. The resultant crop plant is
 CC useful for obtaining a product of a wheat plant, especially grain, which
 CC is optionally processed into flour or another grain product. Food
 CC comprising the antagonistic protein is useful instead of a wild-type
 CC gliadin.

CC Sequence 266 AA;

Query Match 25.8%; Score 234; DB 22; Length 266;

Best Local Similarity 31.8%; Pred. No. 4,2e-13;

Matches 74; Conservative 15; Mismatches 50; Indels 94; Gaps 11;

QY 3 QLNPCSGELGSPQSS---YLDQFYP--QNDYLPQKRFPPYQGFHTPPQYFYLPELEFPQ 57

DB 7 QLDQPNFSQDQPPQDQVPLVQDQDQFQDQDQDQFPPQDQFPPQDQFPPQDQFPPQ 63

QY 58 YQIFPPYLPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 111

DB 64 PQLFYP--QDGS--FP--PQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 118

QY 112 -----PQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 111

DB 119 LQQQILPCMDVYLDQHNIAHARSQVLDQSTYQLDELCCQHLMQILPEQSGCCQAIHNVVHA 178

QY 112 ----PQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQDQFPPQ 149

DB 179 ILHQQQKQDQ 231

RESULT 9

AAU34519
 AAU34519 standard; Protein; 1329 AA.

AC AAU34519;

DT 14-FEB-2002 (first entry)

DE E. coli cellular proliferation protein #100.

KM Antisense: prokaryotic cellular proliferation protein;

OS antibiotic; antibacterial; drug design.

OS Escherichia coli.

PN WO200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WO-US09180.

PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2001; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

DR N-PSDB: AAS52378.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3: Seq ID No 10112; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 1329 AA;

Query Match 24.0%; Score 217; DB 22; Length 1329;

Best Local Similarity 37.3%; Pred. No. 6,7e-11;

Matches 59; Conservative 18; Mismatches 51; Indels 30; Gaps 9;

QY 9 QELQSPQDQSTL-QQPYPNPVL-----PQKPF--PYQDPEHTPQDQFPPY 49

DB 689 QFAQTQDQDQSGEPAGANFSLDDFEFSPKALLDDCPHEPLFTPIVEYDQDQDQ--PV 746

QY 50 LPFELFPQYQIPPLQDQDQFPPQDQDQFPPQDQDQFPPQDQDQFPPQDQDQFPPQDQDQFPPQ 108

DB 747 APQDQ 804

QY 109 PQDQ 141

DB 805 -QQDQ 841

RESULT 10

AAU44499
 AAU44499 standard; Protein; 308 AA.

AC AAU44499;

DT 06-FEB-2001 (first entry)

DE Plant viral movement protein SEQ ID 12.

KM Plant viral movement protein; transport; transgenic plant;

OS viral resistance; cosuppression.

OS Oryza sativa.

PN WO200060088-A2.

XX 12-OCT-2000.

PD 06-APR-2000; 2000WO-US09110.

PF 07-APR-1999; 99US-0128092.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Kriebbers E, Weng Z, Cahoon RE;

PI WPI: 2000-638467/61.

DR N-PSDB: AAC79353.

XX Novel viral movement polypeptides and polynucleotides useful in field

PT of plant molecular biology, for producing transgenic plants, to prepare

PT antibodies and in immunological screening of cDNA expression libraries

XX PS Claim 10; Page 40-41; 62pp; English.
 CC CC Polynucleotide sequences AAC79348-C79375 encode plant viral movement
 CC proteins AAB44494-B44520. Some plant viruses have been shown to be able
 CC to establish systemic infections via movement proteins that utilize
 CC existing plant pathways to traffic macromolecules to surrounding cells.
 CC CC Proteins such as those of the invention are similar to viral movement
 CC proteins that function in the transport of nucleic acids from cell to
 CC cell. The plant viral movement proteins are useful for obtaining a
 CC nucleic acid fragment encoding a viral movement protein. Polynucleotides
 CC encoding the plant viral movement proteins are useful for positive
 CC selection of a transformed cell. The proteins are useful in the field of
 CC plant molecular biology, and in the preparation of antibodies against the
 CC proteins. The proteins are also useful for isolating cDNAs and genes
 CC encoding homologous proteins from the same or other plant species, and to
 CC create transgenic plants in which the protein is presented at higher or
 CC lower levels than normal or in cell types or developmental stages in
 CC which they are not normally found. The proteins and nucleotide sequences
 CC may be used to control cosuppression and engineer plant virus
 CC resistance.
 CC XX Sequence 308 AA:
 SQ
 Query Match 23.5%; Score 213; DB 21; Length 308;
 Best Local Similarity 33.7%; Pred. No. 3.4e-11;
 Matches 57; Conservative 15; Mismatches 63; Indels 34; Gaps 9;
 OY 13 SPQOSTYLOQPPYQNPY-----LPQKFPYVQPF-----HTPOY----FPYLPEEL 54
 Db 140 STHSVLPVPVAVPVAAPSPSYALPPAGYPAVPPYOSYPASHVPAPYPSAHPHPPSL 199
 OY 55 FPO-----YQIFPPLPQPPQPPQPPQ-----QPLPFPQPPFQPPQPPQPPQPPQ-EP 101
 Db 200 LARDVEMAAVPRSTTTPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 258
 OY 102 QOPQ-QPFPQPPQ-QPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 148
 Db 259 PKPQGGPPPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 307
 RESULT 11
 AAM24322
 ID AAM24322 standard; Protein: 1299 AA.
 XX AC AAM24322;
 XX DE 12-OCT-2001 (first entry)
 XX DE Human EST encoded protein SEQ ID NO: 1847.
 XX XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition.
 XX OS Homo sapiens.
 XX PN WO200154477-A2.
 XX PD 02-AUG-2001.
 XX PF 25-JAN-2001; 2001WO-US02687.
 XX PX 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX XX (HYSE-) HYSEQ INC.
 PA Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werthman T;

XX DR WPI; 2001-476164/51.
 DR N-PSDB; AAH98981.
 XX XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX PS Claim 20; Page 1198-1201; 1275pp; English.
 CC CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.
 CC XX Sequence 1299 AA:
 SQ
 Query Match 22.5%; Score 204; DB 22; Length 1299;
 Best Local Similarity 30.2%; Pred. No. 9.1e-10;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
 OY 12 OSPQOSTYLOQPPYQNPYLPQKFPYVQPPYQPPYLPPELPPQYQIPTLPQPPQPP 71
 Db 394 KEPAPTTKKEPAPTTKKEP-APTTTKKEPAPTTKSAPTTKP-----PAPTTPKKPAP 445
 OY 72 QOPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 130
 Db 446 TTREKPAPTTKKEPPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 505
 OY 131 SYVPOQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP 148
 Db 506 PAPTTPKEPPTTKKEPAP 524
 RESULT 12
 AAR26049
 ID AAR26049 standard; Protein: 1404 AA.
 XX AC AAR26049;
 XX DE 02-FEB-1993 (first entry)
 XX DE MSF precursor.
 XX KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
 KW stability; proteolytic cleavage; adhesion; alternative splicing.
 XX OS Synthetic.
 XX XX
 FH key Location/Qualifiers
 FT Region 1..26 /label= Exon_I
 FT Region 26..67 /label= Exon_II
 FT Region 67..107 /label= Exon_III
 FT Region 107..157 /label= Exon_IV
 FT Region 157..200 /label= Exon_V
 FT Region 200..1141 /label= Exon_VI
 FT Region 1141..1166 /label= Exon_VII
 FT Region 1166..1212 /label= Exon_VIII
 FT Region 1213..1266 /label= Exon_IX
 FT Region 1266..1331 /label= Exon_X
 FT Region 1331..1331 /label= Exon_X

FT Region 1331..1373
 FT /label= Exon_XI
 FT Region 1373..1404
 FT /label= Exon_XII
 XX W09213075-A.
 XX
 XX 06-AUG-1992.
 XX
 XX 17-JAN-1992: 92WO-US00433.
 XX
 XX 18-JAN-1991: 91US-0643502.
 PR 10-SEP-1991: 91US-0757022.
 XX
 XX (GENE) GENETICS INST INC.
 XX
 XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K:
 XX
 XX WPI: 1992-284660/34.
 DR N-PSDB: AAQ27223.
 XX
 PT New human mega-karyocyte stimulating factors - for treating
 PT immune deficiencies, cancer, exposure to radiation or drugs,
 PT bacterial and viral infections, etc.
 XX
 PS Claim 1, 2 and 3; Fig 1, 87pp; English.
 XX
 CC The sequence given is a full length translation from the megakaryocyte
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This
 CC sequence is modified by the addition of an N-terminal sequence encoding
 CC a secretory leader, an initiating methionine preceding exon II and a
 CC terminating codon following exon IV. The cDNA sequence given contains
 CC sequences derived from human megakaryocyte colony stimulating factor
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
 CC classical mammalian protein secretion signal sequence. The sequence
 CC encoding the original meg-CSF includes exons II-IV and is thought to
 CC terminate in the region between amino acid residues 134 - 147. The
 CC primary transcript of this gene may be cleaved in different ways to
 CC yield a family of mRNA's each encoding a different MSF protein. Exons
 CC V and VI are thought to be related to the activity of the factor and
 CC are also implicated in the stability, folding and processing of the
 CC molecule. These exons are also thought to play a role in the observed
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be
 CC implicated in the processing or folding of the appropriate structure of
 CC the resulting factor, ie. one or more of these exons may contain
 CC sequences which direct proteolytic cleavage, adhesion, organisation of
 CC the cellular matrix or extracellular matrix processing. Both naturally
 CC occurring and non-naturally occurring MSF's may be characterised by
 CC various combinations of alternatively spliced exons from this sequence,
 CC with the exons spliced together in differing orders to form different
 CC members of the MSF family.
 XX
 XX Sequence 1404 AA:
 SO
 Query Match 22.5%; Score 204; DB 13; Length 1404;
 Best Local Similarity 30.2%; Pred. No. 9.8e-10;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
 QY 12 QSFQOSYLQOQPYPNMPLPQKPFVQOQPFHTPOQYFPLPEELPQOQITPLQOQPPF 71
 Db 394 KEPAPTTKEPAPTTKEP-APTTKEPAPTTKEP-----PAPTTRKKPAP 445
 QY 72 QOQOQPLP-RPOQFPWQOQOQFPPOQEPITQOQOQOQFPPOQOQOQIIFOQOQO 130
 Db 446 TTPEKEPAPTTKEPPTTKEPAPTTPKEPAPTTPKEPAAPAKKPAPTTKEPAPTTKE 505
 QY 131 SYVQOQOQFPPOQ-QPVP 148
 Db 506 PAPTTRKEPSTTPKEPAP 524
 RESULT 13

AAB60568
 ID AAB60568 standard; protein: 1404 AA.
 XX
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 DE Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
 KM MSF: megakaryocyte stimulating factor; synovial lubricant;
 KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 KM antiarthritic.
 XX
 XX Homo sapiens.
 OS
 PN W0200107068-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 21-JUL-2000: 2000WO-US20002.
 XX
 PR 23-JUL-1999: 99US-0145328.
 PR 19-JUL-2000: 2000US-0145328.
 XX
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 PA
 XX Warman ML;
 PI
 DR WPI: 2001-182721/18.
 XX
 PT New composition comprising the campodactylly-arthropathy-coxa
 PT vara-pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints
 XX
 PS Example 1; Page -: 34pp; English.
 XX
 CC The invention relates to a method of treating osteoarthritis via the
 CC administration of a composition comprising the campodactylly-arthropathy-
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 CC The composition may further comprise a local anesthetic. The composition
 CC of the invention may be administered via intra-articular or intravenous
 CC injection. The human CACP protein is identified in the invention as
 CC being megakaryocyte stimulating factor (MSF). The gene encoding
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 CC this gene are responsible for the heritable disorder campodactylly-
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)
 CC acts as a synovium lubricant, and can be used to lubricate tissue and
 CC joints in the treatment of osteoarthritis. The composition may be
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 CC loss of range of movement or joint damage). The present sequence
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
 CC Note: This sequence is not given in its entirety in Figure 4 of the
 CC specification, although a Genbank accession number was given. This
 CC sequence was therefore obtained from Genbank (U070316).
 XX
 XX Sequence 1404 AA:
 SO
 Query Match 22.5%; Score 204; DB 22; Length 1404;
 Best Local Similarity 30.2%; Pred. No. 9.8e-10;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
 QY 12 QSFQOSYLQOQPYPNMPLPQKPFVQOQPFHTPOQYFPLPEELPQOQITPLQOQPPF 71
 Db 394 KEPAPTTKEPAPTTKEP-APTTKEPAPTTKEP-----PAPTTRKKPAP 445
 QY 72 QOQOQPLP-RPOQFPWQOQOQFPPOQEPITQOQOQOQFPPOQOQOQIIFOQOQO 130
 Db 446 TTPEKEPAPTTKEPPTTKEPAPTTPKEPAPTTPKEPAAPAKKPAPTTKEPAPTTKE 505
 QY 131 SYVQOQOQFPPOQ-QPVP 148
 Db 506 PAPTTRKEPSTTPKEPAP 524

DB 506 PAPTTKEPSPPTPKKPAP 524

RESULT 14
AAB29773

ID AAB29773 standard; Protein: 1404 AA.

XX AAB29773;

DT 28-FEB-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF; megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KM osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic;

XX osteopathic.

XX Homo sapiens.

XX WO200064930-A2.

XX 02-NOV-2000.

XX 24-APR-2000; 2000WO-US10953.

XX 23-APR-1999; 9905-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Jay GD;

XX WPI: 2001-024673/03.

XX N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating

XX osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonectin which is a product of

XX alternative splicing of the human MSF (megakaryocyte stimulating factor)

XX gene. The tribonectin has at least one O-linked oligosaccharide

XX lubricating moiety and has a polypeptide sequence comprising 1-76

XX repeats of a motif having at least 50% identity to the sequence KEPAPTT

XX (AAB29774). The invention also relates to a nucleic acid encoding a

XX human MSF-derived tribonectin; a biocompatible composition comprising a

XX human tribonectin for inhibiting tissue adhesion formation; and a method

XX of diagnosing osteoarthritis or a predisposition to osteoarthritis by

XX measuring the amount of MSF or its fragment in a biological sample of a

XX mammal, wherein an increased amount of MSF compared to a control

XX indicates the presence of or predisposition to developing

XX osteoarthritis. The tribonectin and DNA encoding it are useful in the

XX treatment of osteoarthritis, where they may be used for lubricating

XX mammalian joints, such as articulating joints of humans, dogs or horses.

XX The tribonectin, when formulated as a membrane, foam, gel or fibre, is

XX useful for inhibiting adhesion between two surfaces such as the injured

XX tissues of a mammal, where the injury is caused by a surgical insertion

XX or trauma, or an artificial device e.g., an orthopaedic implant. In

XX particular, one of the surfaces is pericardial tissue. DNA encoding a

XX tribonectin may be used in gene therapy. The present sequence represents

XX human MSF.

XX Sequence 1404 AA:

XX

Query Match 22.5%; Score 204; DB 22; Length 1404;

Best Local Similarity 30.2%; Pred. No. 9.8e-10;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQSYLQPPYQNPYLQKRRPVQOQRRHTPOQYFRLPRLPELTPQVQIPIRLQPOQPP 71

DB 394 KEPAPTTKEPAPPTPKRP-APTITKEPAPTTKSAPTPKR-----PAPTPKRPAP 445

OY 72 QDPQQLP-RPQQPPFWQPOQPPFQPPQOQPPQOQPPQOQPPQOQPPQOQPPQO 130
DB 446 TTPKEPAPPTPKRPPTPKRPAPPTPKRPAPPTPKRPAPPTPKRPAPPTPKR 505
OY 131 SYPVQPOQPPPOQ-OPVP 148
DB 506 PAPTTKEPSPPTPKKPAP 524

RESULT 15
AAU32262

ID AAU32262 standard; Protein: 1415 AA.

XX AAU32262;

DT 18-DEC-2001 (first entry)

DE Novel human secreted protein #2753.

XX Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

XX WO200179449-A2.

XX 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

XX 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

XX Tang YF, Liu C, Drmanac RT;

XX WPI: 2001-611725/70.

XX Claim 20; Page 573; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The

XX polypeptides and antibodies to the polypeptides are useful for

XX determining the presence of or predisposition to a disease associated

XX with altered levels of polypeptide. The polypeptides are also useful for

XX identifying agents (agonists and antagonists) that bind to them. Cells

XX expressing the proteins are useful for identifying a therapeutic agent

XX for use in treatment of a pathology related to aberrant expression or

XX physiological interactions of the polypeptide. Vectors comprising

XX the nucleic acids encoding the polypeptides and cells genetically

XX engineered to express them are also useful for producing the proteins.

XX The proteins are useful in genetic vaccination, testing and

XX therapy, and can be used as nutritional supplements. They may be used to

XX increase stem cell proliferation; to regulate haematopoiesis; and in

XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;

XX immune suppression and/or stimulation; as anti-inflammatory agents; and

XX in treatment of leukaemias. AAU29510-AAU33504 represent the amino acid

XX sequences of novel human secreted proteins of the invention.

XX Sequence 1415 AA:

XX

Query Match 22.5%; Score 204; DB 22; Length 1415;

Best Local Similarity 30.2%; Pred. No. 9.9e-10;

Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQSYLQPPYQNPYLQKRRPVQOQRRHTPOQYFRLPRLPELTPQVQIPIRLQPOQPP 71

DB 405 KEPAPTTKEPAPPTPKRP-APTITKEPAPTTKSAPTPKR-----PAPTPKRPAP 456

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:45:16 : Search time 15 Seconds

(without alignments)
309.921 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 1 MR0LNCSCQELQSPQOSTYIQ.....PFPQOPQPVQOASCTSMW 158

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :
1: Issued_patents_AA:*
2: /cgn2_6/p/ptodata/1/iaa/5A.COMB.pep:*
3: /cgn2_6/p/ptodata/1/iaa/5B.COMB.pep:*
4: /cgn2_6/p/ptodata/1/iaa/5A.COMB.pep:*
5: /cgn2_6/p/ptodata/1/iaa/5B.COMB.pep:*
6: /cgn2_6/p/ptodata/1/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	262.5	29.0	1162	2	US-08-728-323A-2	Sequence 2, Appl1
2	262.5	29.0	1162	4	US-09-298-568-2	Sequence 2, Appl1
3	255.5	28.2	941	4	US-08-991-300-2	Sequence 2, Appl1
4	204	22.5	941	4	US-07-757-022B-14	Sequence 14, Appl1
5	204	22.5	1022	4	US-07-757-022B-84	Sequence 84, Appl1
6	204	22.5	1038	4	US-07-757-022B-74	Sequence 74, Appl1
7	204	22.5	1049	4	US-07-757-022B-58	Sequence 58, Appl1
8	204	22.5	1140	4	US-07-757-022B-104	Sequence 104, Appl1
9	204	22.5	1270	4	US-07-757-022B-44	Sequence 44, Appl1
10	204	22.5	1311	4	US-07-757-022B-42	Sequence 42, Appl1
11	204	22.5	1313	4	US-07-757-022B-142	Sequence 142, Appl1
12	204	22.5	1314	4	US-07-757-022B-50	Sequence 50, Appl1
13	204	22.5	1320	4	US-07-757-022B-46	Sequence 46, Appl1
14	204	22.5	1320	4	US-07-757-022B-60	Sequence 60, Appl1
15	204	22.5	1354	4	US-07-757-022B-48	Sequence 48, Appl1
16	204	22.5	1363	4	US-07-757-022B-40	Sequence 40, Appl1
17	204	22.5	1404	4	US-07-757-022B-52	Sequence 52, Appl1
18	204	22.5	1404	4	US-07-757-022B-2	Sequence 2, Appl1
19	204	22.5	1404	4	US-07-757-022B-62	Sequence 62, Appl1
20	184	20.6	1274	4	US-09-095-443-2	Sequence 2, Appl1
21	184	20.6	1677	4	US-08-718-661-2	Sequence 2, Appl1
22	180	19.9	148	4	US-09-178-509-1	Sequence 1, Appl1
23	172	19.0	788	4	US-08-918-914-4	Sequence 4, Appl1
24	169	18.7	605	2	US-08-687-956A-1	Sequence 1, Appl1
25	157.5	17.4	478	3	US-08-153-888-2	Sequence 2, Appl1
26	157.5	17.4	865	4	US-09-281-766-19	Sequence 19, Appl1
27	154	17.0	328	2	US-08-651-818A-2	Sequence 2, Appl1

28	154	17.0	328	4	US-09-184-826-2	Sequence 2, Appl1
29	153	16.9	729	4	US-09-625-188-20	Sequence 20, Appl1
30	150.5	16.6	521	5	PCT-US93-08386-10	Sequence 10, Appl1
31	146.5	16.2	1142	2	US-08-993-118-7	Sequence 7, Appl1
32	146.5	16.2	1142	2	US-08-845-528C-7	Sequence 7, Appl1
33	146.5	16.2	1142	4	US-09-061-709-2	Sequence 2, Appl1
34	146	16.1	2414	5	US-08-227-536-2	Sequence 2, Appl1
35	145.5	16.1	221	4	PCT-US95-04682-2	Sequence 224, App
36	145.5	16.1	282	4	US-09-071-035-224	Sequence 222, App
37	145.5	16.1	312	3	US-08-894-017-10	Sequence 10, Appl1
38	143.5	15.8	1561	3	US-08-894-017-23	Sequence 23, Appl1
39	143.5	15.8	1561	6	US-07-638-431-2	Sequence 2, Appl1
40	143.5	15.8	1565	6	5352450-2	Sequence 2, Appl1
41	142.5	15.7	826	5	PCT-US92-00018-2	Sequence 10, Appl1
42	142.5	15.7	826	5	US-09-281-766-10	Sequence 10, Appl1
43	140.5	15.5	314	2	US-08-525-742-6	Sequence 6, Appl1
44	138	15.2	675	4	US-08-973-273-5	Sequence 5, Appl1
45	137	15.1				

ALIGNMENTS

RESULT 1
US-08-728-323A-2
Sequence 2, Appl1
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 05/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 29.0% Score 262.5 DB 2 Length 1162:
Best local similarity 46.8% Pred. No. 5.2e-18:
Matches 74: Conservative 10: Mismatches 51: Indels 23: Gaps 8:
OY 2 ROLNCSOE-LQSPQOSTYIQPPQNPVLPQKPFQVQPFHTPOQYFPVLPPEEL-FPQYO 59


```

: FILING DATE: 19910910
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/546,114
: FILING DATE: 29-JUN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/457,196
: FILING DATE: 29-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Cserr, Luann
: REGISTRATION NUMBER: 31,822
: REFERENCE/DOCKET NUMBER: GI 5190
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)876-1170
: TELEFAX: (617)876-5851
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 941 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-757-022B-14

Query Match      22.5%; Score 204; DB 4; Length 941;
Best Local Similarity 30.2%; Pred. No. 2.3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOOSYLQOPVQONVYLPQKRFVQOQPFVLPQOYFPLPEELFQYOIPTPLQPOQPP 71
DB 195 KRPAPTTTKRPAPTTKRP-APTTTKRPAPTTKSAPTTTKE-----PAPTTPKRPAP 246
QY 72 QOPQOPLP-RPOQPFVQOQPFPOQPEPIPOQPOQPFPOQPOQPFPOQPOQIIFQOPOQ 130
DB 247 TTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 306
QY 131 SYVPOQOQPFPOQ-PQVP 148
DB 307 PAPTTTKRPSPTTKEPAP 325

RESULT 5
US-07-757-022B-84
: Sequence 84, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/757,022B
: FILING DATE: 19910910
: CLASSIFICATION: 530
```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/546,114
: FILING DATE: 29-JUN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/457,196
: FILING DATE: 29-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Cserr, Luann
: REGISTRATION NUMBER: 31,822
: REFERENCE/DOCKET NUMBER: GI 5190
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)876-1170
: TELEFAX: (617)876-5851
: INFORMATION FOR SEQ ID NO: 84:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1022 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-757-022B-84

Query Match      22.5%; Score 204; DB 4; Length 1022;
Best Local Similarity 30.2%; Pred. No. 2.5e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSPOOSYLQOPVQONVYLPQKRFVQOQPFVLPQOYFPLPEELFQYOIPTPLQPOQPP 71
DB 276 KRPAPTTTKRPAPTTKRP-APTTTKRPAPTTKSAPTTTKE-----PAPTTPKRPAP 327
QY 72 QOPQOPLP-RPOQPFVQOQPFPOQPEPIPOQPOQPFPOQPOQPFPOQPOQIIFQOPOQ 130
DB 328 TTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 387
QY 131 SYVPOQOQPFPOQ-PQVP 148
DB 388 PAPTTTKRPSPTTKEPAP 406

RESULT 6
US-07-757-022B-74
: Sequence 74, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/757,022B
: FILING DATE: 19910910
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
```

;; FILING DATE: 18-JAN-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cseerr, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)876-1170
;; TELEFAX: (617)876-5851
;; INFORMATION FOR SEQ ID NO: 74:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1038 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 22.5%; Score 204; DB 4; Length 1038;
Best Local Similarity 30.2%; Pred. No. 2.6e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQOPYPONPYLPQKPEPVQOPHTPQOYFPYLPBELFPYOIPTPLQPOQPP 71
DB 260 KEPAATTTKEPAATTTKEP-APTITKEPAATTTKSAPTTPKE-----PAPTTPKKPAP 311
QY 72 QOPQOPLP-RPOQFPWQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQ 130
DB 312 TTPKRPATTTKEPTTTPKPEPATTTKEPATTTKEPATTTAKKAPATTTKEPATTTTKE 130
QY 131 STPVQOPQOPQOP-QPVP 148
DB 372 PAPTTTKEPSPTTKEPAP 390

RESULT 7
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cseerr, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)876-1170
;; TELEFAX: (617)876-5851
;; INFORMATION FOR SEQ ID NO: 58:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1049 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-757-022B-58

Query Match 22.5%; Score 204; DB 4; Length 1049;
Best Local Similarity 30.2%; Pred. No. 2.6e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQOPYPONPYLPQKPEPVQOPHTPQOYFPYLPBELFPYOIPTPLQPOQPP 71
DB 303 KEPAATTTKEPAATTTKEP-APTITKEPAATTTKSAPTTPKE-----PAPTTPKKPAP 354
QY 72 QOPQOPLP-RPOQFPWQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQOPQ 130
DB 355 TTPKRPATTTKEPTTTPKPEPATTTKEPATTTKEPATTTAKKAPATTTKEPATTTTKE 414
QY 131 STPVQOPQOPQOP-QPVP 148
DB 415 PAPTTTKEPSPTTKEPAP 433

RESULT 8
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-104

Query Match 22.5%; Score 204; DB 4; Length 1140;
Best Local Similarity 30.2%; Pred. No. 2.9e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4:

QY 12 QSPQOSYLQOPYPQNPYLPQKPFVQOPEHTPQOYFPLPEELFPQYQIPTPLQPOQPP 71
DB 394 KEFAPPTTKKEPAPPTTKKEP-APTTKKEPAPPTTKSAPTTKKE-----PAPTPPKKRPAP 445
QY 72 QQPQOPLP-RPQOPFPWQOPFPQOPPEPTPQOPQOPFPQOPFPQOPFPQOQIIFQOPOQ 130
DB 446 TTPKEAPPTTKKEPPTTKKEPAPPTTKKEPAPPTTKKEPAPAPKRPAPPTTKKEPAPPTPK 505
QY 131 SYVPOQOPFPQOP-QPVP 148
DB 506 PAPTTTKESPPTTKKEPAP 524

RESULT 9
US-07-757-022B-44
Sequence 44, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-44

Query Match 22.5%; Score 204; DB 4; Length 1270;
Best Local Similarity 30.2%; Pred. No. 3.2e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4:

QY 12 QSPQOSYLQOPYPQNPYLPQKPFVQOPEHTPQOYFPLPEELFPQYQIPTPLQPOQPP 71
DB 260 KEFAPPTTKKEPAPPTTKKEP-APTTKKEPAPPTTKSAPTTKKE-----PAPTPPKKRPAP 311
QY 72 QQPQOPLP-RPQOPFPWQOPFPQOPPEPTPQOPQOPFPQOPFPQOPFPQOQIIFQOPOQ 130
DB 312 TTPKEAPPTTKKEPPTTKKEPAPPTTKKEPAPPTTKKEPAPAPKRPAPPTTKKEPAPPTPK 371
QY 131 SYVPOQOPFPQOP-QPVP 148
DB 372 PAPTTTKESPPTTKKEPAP 390

RESULT 10
US-07-757-022B-42
Sequence 42, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

```
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-42

Query Match          22.5%; Score 204; DB 4; Length 1311;
Best Local Similarity 30.2%; Pred. No. 3.3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQDPYQNPYLPOKFPVQOPFPHTPOQYFPLYLPELFPYQYQIPPLQPOOPFP 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 KEPAPTTKKEPAPTTKKEP-APTTKKEPAPTTKSAPTTKPE-----PAPTTPKKPAP 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 QQPQOPLP-RPQOPFWOQOPFPPOPEPIPOQPOOPFPPOQPOQPIIFQOPOQ 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 353 TTKKEPAPTTKKEPTTKKEPAPTTKKEPAPTTKKEPAAPAKKPAPTTKKEPAPTTKPE 412
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 SYPVQOOPFPPOOP-OPVP 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 413 PAPTTKKEPSPTTKKEPAP 431

RESULT 11
US-07-757-022B-142
: Sequence 142, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/757,022B
: FILING DATE: 19910910
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/546,114
: FILING DATE: 29-JUN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/457,196
: FILING DATE: 29-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Cseert, Luann
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:
SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-142

Query Match          22.5%; Score 204; DB 4; Length 1313;
Best Local Similarity 30.2%; Pred. No. 3.3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQDPYQNPYLPOKFPVQOPFPHTPOQYFPLYLPELFPYQYQIPPLQPOOPFP 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 KEPAPTTKKEPAPTTKKEP-APTTKKEPAPTTKSAPTTKPE-----PAPTTPKKPAP 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 QQPQOPLP-RPQOPFWOQOPFPPOPEPIPOQPOOPFPPOQPOQPIIFQOPOQ 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 355 TTKKEPAPTTKKEPTTKKEPAPTTKKEPAPTTKKEPAAPAKKPAPTTKKEPAPTTKPE 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 SYPVQOOPFPPOOP-OPVP 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 415 PAPTTKKEPSPTTKKEPAP 433

RESULT 12
US-07-757-022B-50
: Sequence 50, Application US/07757022B
: Patent No. 6433142
: GENERAL INFORMATION:
: APPLICANT: Gesner, Thomas G.
: APPLICANT: Clark, Stephen C.
: APPLICANT: Turner, Katherine
: APPLICANT: Hewick, Rodney M.
: TITLE OF INVENTION: Megakaryocyte Stimulating Factors
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/757,022B
: FILING DATE: 19910910
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/643,502
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/546,114
: FILING DATE: 29-JUN-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/457,196
: FILING DATE: 29-DEC-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Cseert, Luann
```

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 22.5%; Score 204; DB 4; Length 1314;
Best Local Similarity 30.2%; Pred. No. 3,3e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
QY 12 QSPQSYLQDPYQNPYLPQKPPVQOQPFPHPOQYFPLYLPELFPQYQIPIPLQOQOQFP 71
DB 304 KEPAPTTKKEPAPTTKPEP-APTTKKEPAPTTKSAPTTKPE-----PAPTTPKKKAP 355
QY 72 QOQOQPLP-RPOQPFPMQOQPFPOQEP1POQPOQPFQOQOQPFQOQOQIIFQOQPOQ 130
DB 356 TTPKEPAPTTKKEPTTTPKEPAPTTKKEPAPTTKKEPAATAKKAAPTTPKEPAPTTKPE 415
QY 131 SYPVQPOQPFPOQF-QVP 148
DB 416 PAPTTPKEPPTTKKEPA 434

RESULT 13
US-07-757-022B-46
Sequence 46, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-46

Query Match 22.5%; Score 204; DB 4; Length 1320;
Best Local Similarity 30.2%; Pred. No. 3,4e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;
QY 12 QSPQSYLQDPYQNPYLPQKPPVQOQPFPHPOQYFPLYLPELFPQYQIPIPLQOQOQFP 71
DB 310 KEPAPTTKKEPAPTTKPEP-APTTKKEPAPTTKSAPTTKPE-----PAPTTPKKKAP 361
QY 72 QOQOQPLP-RPOQPFPMQOQPFPOQEP1POQPOQPFQOQOQPFQOQOQIIFQOQPOQ 130
DB 362 TTPKEPAPTTKKEPTTTPKEPAPTTKKEPAPTTKKEPAATAKKAAPTTPKEPAPTTKPE 421
QY 131 SYPVQPOQPFPOQF-QVP 148
DB 422 PAPTTPKEPPTTKKEPA 440

RESULT 14
US-07-757-022B-60
Sequence 60, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170

Wed May 28 09:17:52 2003

us-09-743-533-19.ra1

Page 8

```
TELEFAX: (617)876-5851
:
: INFORMATION FOR SEQ ID NO: 60:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1320 amino acids
:   TYPE: AMINO ACID
:   TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: us-07-757-022B-60

Query Match      22.5%; Score 204; DB 4; Length 1320;
Best Local Similarity 30.2%; Pred. No. 3.4e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 OSPOQSYLQOPYPQNPVLPQKPFVQOPFHTPQGYFPLPEELFPOYQITPLQPOQPP 71
DB 310 KEPAFTTKKEPAFTTPKEP-APTITKEPAFTTKSAFTTPKE-----PAFTTPKKEPAP 361
OY 72 QOPQOPLP-RPOQPFVQOPQOPFPOQEPPIPOQOPQOPQOPQOPQOIIIFQOPOQ 130
DB 362 TTKKEPAFTTPKEPPTTPKEPAFTTPKEPAFTTPKEPAFTTPKEPAFTTPKE 421
OY 131 SYVQOPQOPFPOQ-OPVP 148
DB 422 PAFTTKKEPSPTTPKEPAP 440

RESULT 15
US-07-757-022B-48
: Sequence 48, Application US/07757022B
: Patent No. 6433142
:
: GENERAL INFORMATION:
:   APPLICANT: Gesner, Thomas G.
:   APPLICANT: Clark, Stephen C.
:   APPLICANT: Turner, Katherine
:   APPLICANT: Hewick, Rodney M.
:   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
:   NUMBER OF SEQUENCES: 143
:   CORRESPONDENCE ADDRESS:
:     ADDRESS: Genetics Institute, Inc.
:     STREET: 87 Cambridgepark Drive
:     CITY: Cambridge
:     STATE: Massachusetts
:     COUNTRY: U.S.A.
:     ZIP: 02140
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentia Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/07/757,022B
:   FILING DATE: 19910910
:   CLASSIFICATION: 530
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 07/643,502
:     FILING DATE: 18-JAN-1991
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 07/546,114
:     FILING DATE: 29-JUN-1990
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 07/457,196
:     FILING DATE: 29-DEC-1989
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: US 07/390,901
:     FILING DATE: 08-AUG-1989
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Casert, Luann
:     REGISTRATION NUMBER: 31,822
:     REFERENCE/DOCKET NUMBER: GI 5190
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (617)876-1170
:       TELEFAX: (617)876-5851
:
: INFORMATION FOR SEQ ID NO: 48:
```

```
SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1354 amino acids
:
: TYPE: AMINO ACID
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: us-07-757-022B-48
```

```
Query Match      22.5%; Score 204; DB 4; Length 1354;
Best Local Similarity 30.2%; Pred. No. 3.5e-12;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 OSPOQSYLQOPYPQNPVLPQKPFVQOPFHTPQGYFPLPEELFPOYQITPLQPOQPP 71
DB 344 KEPAFTTKKEPAFTTPKEP-APTITKEPAFTTKSAFTTPKE-----PAFTTPKKEPAP 365
OY 72 QOPQOPLP-RPOQPFVQOPQOPFPOQEPPIPOQOPQOPQOPQOIIIFQOPOQ 130
DB 396 TTKKEPAFTTPKEPPTTPKEPAFTTPKEPAFTTPKEPAFTTPKEPAFTTPKE 455
OY 131 SYVQOPQOPFPOQ-OPVP 148
DB 456 PAFTTKKEPSPTTPKEPAP 474
```

Search completed: May 27, 2003, 14:48:11
Job time : 26 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:45:55 : Search time 21 Seconds

(without alignments)
746.156 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906
Sequence: I MRQLNPGSQRSLSPQDSYU.....PPPOQPPVPOQASCTSMV 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-AA:
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppa/PTCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244.5	27.0	1362	10	US-09-815-242-14009
2	222	24.5	296	9	US-09-789-054A-10
3	217	24.0	1329	10	US-09-815-242-10112
4	204	22.5	941	12	US-10-124-557-14
5	204	22.5	1022	12	US-10-124-557-84
6	204	22.5	1038	12	US-10-124-557-74
7	204	22.5	1049	12	US-10-124-557-58
8	204	22.5	1140	12	US-10-124-557-104
9	204	22.5	1270	12	US-10-124-557-44
10	204	22.5	1311	12	US-10-124-557-12
11	204	22.5	1313	12	US-10-124-557-142
12	204	22.5	1314	12	US-10-124-557-50
13	204	22.5	1320	12	US-10-124-557-46
14	204	22.5	1320	12	US-10-124-557-60
15	204	22.5	1334	12	US-10-124-557-48
16	204	22.5	1361	12	US-10-124-557-40
17	204	22.5	1363	12	US-10-124-557-52
18	204	22.5	1404	12	US-10-124-557-2
19	204	22.5	1404	12	US-10-124-557-62

20	189	20.9	312	9	US-09-789-054A-22	Sequence 22, App1
21	187	20.6	1274	9	US-10-020-215-2	Sequence 2, App1
22	183.5	20.3	371	9	US-09-284-320-2	Sequence 2, App1
23	183.5	20.3	371	9	US-10-197-666A-4	Sequence 4, App1
24	180	19.9	148	12	US-10-117-121-1	Sequence 1, App1
25	171.5	18.9	254	10	US-09-216-393-126	Sequence 126, App
26	167.5	18.5	171	10	US-09-216-393-71	Sequence 71, App1
27	167.5	18.5	171	10	US-09-216-393-274	Sequence 274, App
28	164.5	18.2	171	9	US-10-077-584-4	Sequence 4, App1
29	163.5	18.0	503	12	US-10-062-254-240	Sequence 240, App
30	161.5	17.8	340	9	US-09-789-054A-6	Sequence 6, App1
31	157.5	17.4	159	9	US-10-034-934-125	Sequence 125, App
32	157.5	17.4	865	9	US-09-957-995A-19	Sequence 19, App1
33	157	17.3	348	9	US-09-789-054A-20	Sequence 20, App1
34	156.5	17.3	2005	10	US-09-735-367B-3	Sequence 3, App1
35	156.5	17.3	2063	10	US-09-735-367B-2	Sequence 2, App1
36	156	17.2	370	9	US-10-021-955-78	Sequence 78, App1
37	156	17.2	370	9	US-10-021-955-83	Sequence 83, App1
38	156	17.2	370	9	US-10-021-955-90	Sequence 90, App1
39	155.5	17.0	370	9	US-10-021-955-91	Sequence 91, App1
40	154	17.0	407	9	US-09-738-626-5433	Sequence 5433, Ap
41	153	16.9	666	10	US-09-791-171-70	Sequence 70, App1
42	152	16.8	668	10	US-09-925-300-1633	Sequence 1633, Ap
43	151	16.7	868	10	US-09-800-729-106	Sequence 106, App
44	151	16.7	921	10	US-09-800-729-199	Sequence 199, App
45	150.5	16.6	356	10	US-09-735-367B-13	Sequence 13, App1

ALIGNMENTS

RESULT 1
US-09-815-242-14009
Sequence 14009 Application US/09815242
Patient No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl F.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.01A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 14009
LENGTH: 1362
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(1362)
OTHER INFORMATION: Xaa = Any Amino Acid

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-10-124-557-14

Query Match 22.5%; Score 204; DB 12; Length 941;
Best Local Similarity 30.2%; Pred. No. 6e-08;
Matches 42; Conservative 26; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQPPYLPQKFPVQOQPFHPPQYFLPEELFPQYQIPTPLQOQPP 71
DB 195 KEPAPTTKEPAPTPKPE-APTTPKEPAPTTKSAPTTPKE-----PAPTTPKKPAP 246
QY 72 QGQOQPLP-RPOQFPWQPOQFPQOEPITQOQPOQFPQOQFPQOQIIFQOQ 130
DB 247 TTKEPAPTTKEPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 306
QY 131 STPVQPOQFPQOP-QVP 148
DB 307 PAPTTKEPSPPTTPKEPAP 325

RESULT 5

US-10-124-557-84
Sequence 84, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:

US-10-124-557-84

Query Match 22.5%; Score 204; DB 12; Length 1022;
Best Local Similarity 30.2%; Pred. No. 6.5e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 QSPQSYLQPPYLPQKFPVQOQPFHPPQYFLPEELFPQYQIPTPLQOQPP 71
DB 276 KEPAPTTKEPAPTPKPE-APTTPKEPAPTTKSAPTTPKE-----PAPTTPKKPAP 327
QY 72 QGQOQPLP-RPOQFPWQPOQFPQOEPITQOQPOQFPQOQFPQOQIIFQOQ 130
DB 328 TTKEPAPTTTKEPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 387
QY 131 STPVQPOQFPQOP-QVP 148
DB 388 PAPTTKEPSPPTTPKEPAP 406

RESULT 6

US-10-124-557-74
Sequence 74, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseff, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 22.5% Score 204; DB 12: Length 1038;
Best Local Similarity 30.2%; Pred No. 6.6e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQOSTIQQPYQNPYLPOKRFVQOQPHHTPOQYFPLPELFPYOIPTPLQPOQPP 71
DB 260 KEPAPTTKKEPAPTTKPEP-APTITKEPAPTTKSAPTTKPE-----PAPTTPKKPAP 311
OY 72 QOPOQPLP-RPOQFFPMQPOQPFPOQEPPIPOQPOQPFPOQPOQPFPOQPOQIIFQOPOQ 130
DB 312 TTPKEPAPTTKKEPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 371
OY 131 SYVQPOQPPPOQ-OPVP 148
DB 372 PAPTTPKEPSPTTKPEPAP 390

RESULT 7
US-10-124-557-58
Sequence 58, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseff, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 22.5% Score 204; DB 12: Length 1049;
Best Local Similarity 30.2%; Pred No. 6.7e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 12 QSPQOSTIQQPYQNPYLPOKRFVQOQPHHTPOQYFPLPELFPYOIPTPLQPOQPP 71
DB 303 KEPAPTTKKEPAPTTKPEP-APTITKEPAPTTKSAPTTKPE-----PAPTTPKKPAP 354
OY 72 QOPOQPLP-RPOQFFPMQPOQPFPOQEPPIPOQPOQPFPOQPOQPFPOQPOQIIFQOPOQ 130
DB 355 TTPKEPAPTTKKEPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 414
OY 131 SYVQPOQPPPOQ-OPVP 148
DB 415 PAPTTPKEPSPTTKPEPAP 433

RESULT 8
US-10-124-557-104
Sequence 104, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseff, Luann

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match
Best Local Similarity 30.2%; Pred. No. 7,2e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

22.5%; Score 204; DB 12; Length 1140;
DB 12 QSPQSYLQQYPQNPVLPQKFPVQOPFHTPQOYFPLPEELFPQYQIFPPLQPOQPP 71
DB 394 KEPAPTTKEPAPTTKEP-APTTEKPAPTTTSAPTPE-----PAPTTKKRAP 445
QY 72 QOPOQLP-RPOQPFQWQOPFPOPEPIPOQPOQPOQPOQPOQPOQPOQIIFQOPOQ 130
DB 446 TTKKEPAPTTKEPPTTTPKEPAPTTKEPAPTTKEPAPTTAPKKAPATTKEPAPTTPE 505
QY 131 SYVQPOQPPPOQ-OPVP 148
DB 506 PAPTTKEPSPTTKEPAP 524

RESULT 9
US-10-124-557-44
Sequence 44, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
CLASSIFICATION: <Unknown>
FILING DATE: 16-Apr-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserf, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match
Best Local Similarity 30.2%; Pred. No. 8e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

22.5%; Score 204; DB 12; Length 1270;
DB 260 KEPAPTTKEPAPTTKEP-APTTEKPAPTTTSAPTPE-----PAPTTKKRAP 311
QY 72 QOPOQLP-RPOQPFQWQOPFPOPEPIPOQPOQPOQPOQPOQPOQIIFQOPOQ 130
DB 312 TTKKEPAPTTKEPPTTTPKEPAPTTKEPAPTTKEPAPTTAPKKAPATTKEPAPTTPE 371
QY 131 SYVQPOQPPPOQ-OPVP 148
DB 372 PAPTTKEPSPTTKEPAP 390

RESULT 10
US-10-124-557-42
Sequence 42, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
CLASSIFICATION: <Unknown>
FILING DATE: 16-Apr-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserf, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 22.5%: Score 204; DB 12; Length 1311;
Best Local Similarity 30.2%; Pred. No. 8.3e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSQOSYIQOQYPPONVLPQKPPVQOQPPHTPQOQFYPLDELFPQYQIPLQPOQPP 71
DB 301 KEPAFTTKKEPAPTTKPEP-APTITKEPAPTTTKSAPTITKE-----PAPTTKRPAP 352
QY 72 QOQOQPLP-RPOQPFPMQOQPFPOQEPITPOQPOQPFPOQPOQPPQOQIIFQOQPO 130
DB 353 TTPKEPAPTTKPEPPTTPKPEPAPTTKPEPAPTTKPEPAPTAAPKRPAPTTKPEPAPTTTKE 412
QY 131 SYVQPOQOQPFPOQ-OPVP 148
DB 413 PAPTTTKEPSPTTKPEPAP 431

RESULT 11

US-10-124-557-142
; Sequence 142, Application US/10124557
; Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Casert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match 22.5%: Score 204; DB 12; Length 1313;
Best Local Similarity 30.2%; Pred. No. 8.3e-08;
Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

QY 12 OSQOSYIQOQYPPONVLPQKPPVQOQPPHTPQOQFYPLDELFPQYQIPLQPOQPP 71
DB 303 KEPAFTTKKEPAPTTKPEP-APTITKEPAPTTTKSAPTITKE-----PAPTTKRPAP 354
QY 72 QOQOQPLP-RPOQPFPMQOQPFPOQEPITPOQPOQPFPOQPOQPPQOQIIFQOQPO 130
DB 355 TTPKEPAPTTKPEPPTTPKPEPAPTTKPEPAPTTKPEPAPTAAPKRPAPTTKPEPAPTTTKE 414
QY 131 SYVQPOQOQPFPOQ-OPVP 148
DB 415 PAPTTTKEPSPTTKPEPAP 433

RESULT 12

US-10-124-557-50
; Sequence 50, Application US/10124557
; Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Casert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:

OY 12 OSPOSYLOQPPONFYLPQKPFVQOPFHTPQOYFFYLEBELFPQYQIPPLQPOQPP 71
 Db 310 KEPAFTTKKEPAFTTKKE-APTTTKEPAFTTKSAFTTKKE-----PAPTPKPPAP 361
 OY 72 OQPOQPLP-RPOQPFPMQPOQPFPOPEPIPOQPOQFPPOQPOQPOQOIIIFQOPQ 130
 Db 362 TTKEPAFTTKKEPTTTPKEPAFTTKKEPAFTTKKEPAFTAPKPPAFTTKKEPAFTTKKE 421
 OY 131 SYVQPOQFPPOQ-OPVP 148
 Db 422 PAPTTKEPSPTTKKEPA 440

RESULT 15

US-10-124-557-48
 : Sequence 48, Application US/10124557
 : Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1354 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 48:
 US-10-124-557-48

Query Match 22.5%; Score 204; DB 12; Length 1354;
 Best Local Similarity 30.2%; Pred. No. 8.5e-08;
 Matches 42; Conservative 28; Mismatches 59; Indels 10; Gaps 4;

OY 72 OQPOQPLP-RPOQPFPMQPOQPFPOPEPIPOQPOQFPPOQPOQPOQOIIIFQOPQ 130
 Db 396 TTKEPAFTTKKEPTTTPKEPAFTTKKEPAFTTKKEPAFTAPKPPAFTTKKEPAFTTKKE 455
 OY 131 SYVQPOQFPPOQ-OPVP 148
 Db 456 PAPTTKEPSPTTKKEPA 474

Search completed: May 27, 2003, 14:48:34
 Job time : 27 secs

OY 12 OSPOSYLOQPPONFYLPQKPFVQOPFHTPQOYFFYLEBELFPQYQIPPLQPOQPP 71
 Db 344 KEPAFTTKKEPAFTTKKE-APTTTKEPAFTTKSAFTTKKE-----PAPTPKPPAP 395

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:36:30 : Search time 15 Seconds
(without alignments)

436.884 Million cell updates/sec

Title: US-09-743-533-19

Perfect score: 906
Sequence: 1 MROLNPGSGLQSPQOSTYLO.....PFPQOPQVPOQASCTMSW 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412	45.5	327	1	GDB2_WHEAT
2	384.5	42.4	302	1	GDB2_WHEAT
3	347	38.3	251	1	GDB0_WHEAT
4	332	36.6	291	1	GDB8_WHEAT
5	310.5	34.3	105	1	HOR1_HORVU
6	302	33.3	305	1	HOG1_HORVU
7	289.5	32.0	307	1	GDA9_WHEAT
8	272.5	30.1	72	1	HOR8_HORVU
9	271.5	30.0	293	1	HOR1_HORVU
10	267.5	29.5	289	1	HOG3_HORVU
11	263.5	29.1	297	1	GDA4_WHEAT
12	255	28.1	296	1	GDA6_WHEAT
13	250	27.6	291	1	GDA2_WHEAT
14	242	26.7	282	1	GDA3_WHEAT
15	240	26.5	264	1	HOR3_HORVU
16	238	26.3	356	1	GDA7_WHEAT
17	236.5	26.1	319	1	GDA7_WHEAT
18	236	26.0	313	1	GDA7_WHEAT
19	234	25.8	286	1	GDA0_WHEAT
20	232.5	25.7	262	1	GDA1_WHEAT
21	218.5	24.1	68	1	HOR9_HORVU
22	217	24.0	1329	1	FTSB_ECOLI
23	211.5	23.3	307	1	GLTB_WHEAT
24	208.5	23.0	462	1	ANX7_DICDI
25	206	22.7	213	1	AMEX_BOVIN
26	205	22.6	202	1	AMEX_MONDO
27	202	22.3	1902	1	SMEL_HUMAN
28	201.5	22.2	152	1	CORA_RAT
29	196	21.6	467	1	PCPA_DICDI
30	195	21.5	2364	1	PCGA_BOVIN
31	189	20.9	126	1	CORA_RABIT
32	189	20.9	1362	1	BRD4_HUMAN
33	188.5	20.8	144	1	CORA_MOUSE

34	186	20.5	191	1	AMEX_HUMAN
35	182.5	20.1	555	1	Q9F217 rat
36	182	20.1	738	1	YKFA_YEAST
37	181.3	20.0	634	1	HMP1_CANAL
38	181	20.0	255	1	LP61_ETMTE
39	180	19.9	1048	1	SR44_RAT
40	177.5	19.6	196	1	AMEL_MOUSE
41	177	19.5	304	1	GDB1_WHEAT
42	176	19.4	746	1	PCAP_HUMAN
43	175.5	19.4	875	1	Y066_NPYOP
44	175	19.3	3164	1	TECU_HSV11
45	170.5	18.8	244	1	GDB3_WHEAT

ALIGNMENTS

```

RESULT 1
GDB2_WHEAT
ID GDB2_WHEAT STANDARD; PRT: 327 AA.
AC P08453:
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Gamma-gliadin precursor.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiyama T., Rafalski A., Soell D.;
RT "The nucleotide sequence of a wheat gamma-gliadin genomic clone.";
RL Plant Sci. 44:205-209(1986).
CC -i- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -i- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M1064; AAA34289.1; .
CC PIR: JS0402; JS0402.
CC InterPro: IPR003612; AAI.
CC InterPro: IPR001954; G1a-glutenin.
CC InterPro: IPR001768; Try/amyL inhbr.
CC Pfam: PF00234; tryP_alpha_aml; 1.
CC PRINTS: PR00206; GLIADGLUTEN.
CC SMART: SM00499; AAI; 1.
CC Seed storage protein; Repeat; Signal; Multigene family.
CC STGNL 1 327 GAMMA-GLIADIN.
CC FT CHAIN 20 327
CC SO SEQUENCE 327 AA: 37122 MW: E27FE9DABDFCB CRC64;

Query Match 45.5%; Score 412; DB 1; Length 327;
Best local Similarity 50.0%; Pred. No. 3.9e-22;
Matches 97; Conservative 12; Mismatches 37; Indels 48; Gaps 10;
QY 3 QLNCSGE--LQSPQSYLQOPYPQNPYLPRKFPYQOPFPQYFPYLPBELPFOYQI 60
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
22 QVDSGGVQWLDQQLVQQLQPLSQP-----QQTFFQYQYFFPHQPPQGVVPOQ 72
QY 61 PTP--LQPPQFPQPPQPPPLP--RPQOPFPWOPQOPFP--OPQEFIPQOPQOPFP--Q 110
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
73 PQQFFLPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 132

```

QY 111 OPOOPPP--OQPOOIFPOOPQOS--YPOPOOPPOPOPPPOQA----- 151
 Db 133 QPOOPFPQOQQPOQ--FPPOPOQOLPOPOQPOQOSFPQOQNFPIQSLQOOLNPKNILLQO 191
 QY 152 -----SCISMW 158
 Db 192 SKPASTVSLMSIT 205

RESULT 2

GDBX_WHEAT
 ID GDBX_WHEAT STANDARD: PRT: 302 AA.

AC P21292;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE Gamma-gliadin precursor
 OS Triticum aestivum (wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN=cy. Yamhill;
 RA Scheets K., Hedgcock C.;
 RT "Nucleotide sequence of a gamma gliadin gene: comparisons with other
 RT gamma gliadin sequences show the structure of gamma gliadin genes and
 RT the general primary structure of gamma gliadins."
 RL Plant Sci. 57:141-150(1988).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M36999; AAA34272.1; -
 CC PIR: JAO153; JAO153.
 CC InterPro: IPR003612; AAI.
 CC InterPro: IPR001954; Glia-glutenin.
 CC InterPro: IPR001768; Try/amyL_inhbr.
 CC Pfam: PF00234; try_alpha_aml; 1.
 CC PRINTS: PR00208; GLIADGLUTEN.
 CC SMART: SM00499; AAI; 1.
 CC Seed storage protein; Repeat; Signal; Multigene family.
 KW SIGNAL
 FT CHAIN 1 19
 FT CHAIN 20 302
 SO SEQUENCE 302 AA; 34300 MW; 672437C770D6F782 CRC64;

Query Match 42.4%; Score 384.5; DB 1; Length 302;
 Best Local Similarity 58.0%; Pred. No. 2.8e-20;
 Matches 91; Conservative 11; Mismatches 28; Indels 27; Gaps 11;

QY 3 QLNPCSOELSPQOSYLOQPYPONPYLPQKFPFVQOPFHT---POQYFPLBELFPOYQ 59
 Db 22 QVDP-SQGVQWMPQ-----QQPFQ-----PQPF-CQQPGRTIQRPHQTHHQPQOTFFPQ 71
 QY 60 IPIPLQOQPP--OQPOPLPRPOQPFPMQOQPP--OPEPT--POQPOPP--OQP 112
 Db 72 QTYPHQOQOQFPOQOQFFPOQOQTFPOQOLPRPOQOQFFPOQOQOQFFPOSOOP 131
 QY 113 QQPFPOQOQIFPOQPOQSFVPOQOPFPOQOPVPO 149
 Db 132 QQPFPOQOQ--FPQFQ-----QPOQSFPOQOQPAIQ 161

RESULT 3

GDBO_WHEAT
 ID GDBO_WHEAT STANDARD: PRT: 251 AA.

AC P08079;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-AUG-1990 (Rel. 16, Last annotation update)
 DE Gamma-gliadin precursor (Fragment).
 OS Triticum aestivum (wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Scheets K., Rafalski J.A., Hedgcock C., Soell D.G.;
 RA "Heptapeptide repeat structure of a wheat gamma-gliadin."
 RT Plant Sci. Lett. 37:221-225(1985).
 RL Plant Sci. Lett. 37:221-225(1985).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -1- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M10660; AAA34288.1; -
 CC PIR: PS0094; PS0094.
 CC InterPro: IPR001768; Try/amyL_inhbr.
 CC Pfam: PF00234; try_alpha_aml; 1.
 CC Seed storage protein; Repeat; Signal; Multigene family.
 KW SIGNAL
 FT CHAIN 1 19
 FT CHAIN 20 >251
 FT NON-TER 251 251
 SO SEQUENCE 251 AA; 29054 MW; 5A3F12C36C825EAD CRC64;

Query Match 38.3%; Score 347; DB 1; Length 251;
 Best Local Similarity 45.0%; Pred. No. 8.7e-18;
 Matches 86; Conservative 11; Mismatches 24; Indels 70; Gaps 11;

QY 3 QLNPCSOELSPQOSYLOQPYPONPYLPQKFPFVQOPFHT---POQYFPLBELFPOYQ 62
 Db 22 QVDPSSQ-VQWMPQ-----QQPFQ-----FHQPFSSQ-----PQPF 53
 QY 63 PIQOPFPPOQOQOLPRPOQOQFFPMQOQOPFPOQOPEPTLPQOQOQFFPOQOQPP--OQP 120
 Db 54 --OPOQTFPHQOQOQFPQO-----QPOQOFLQOQOQFFPOQOQOQOQOQOQOQOQOQ 106
 QY 121 QQIIT-----FOQPOQSY--PYQPOQPFPOQOQ-----QVPOQO----- 150
 Db 107 QQLFPOQ 166
 QY 151 --ASCISMW 158
 Db 167 VSLVSLMSMI 177

RESULT 4

GDBR_WHEAT
 ID GDBR_WHEAT STANDARD: PRT: 291 AA.

AC P06659;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Gamma-gliadin B precursor.

OS *Triticum aestivum* (Wheat).
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 CC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae:
 CC Triticeae: Triticum.
 CC NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301876; PubMed=3017812;
 RA Rafalski J.A.;
 RT "Structure of wheat gamma-gliadin genes";
 CC Structure of wheat gamma-gliadin genes";
 CC RT gene 43:221-229(1986).
 CC -I- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -I- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13713; AAA34274.1;
 CC PIR: A25632; EEWG.
 CC InterPro: IPR003612; ANI.
 CC InterPro: IPR001954; G1a_gluTenin.
 CC InterPro: IPR001768; try/amy1_inhbr.
 CC Pfam: PF00234; tryP-alpha_amy1.
 CC PRINTS: PR00208; GLIADGLUTEN.
 CC SMART: SM00499; AAI: 1.
 CC Seed storage protein: Repeat; signal; Multigene family.
 CC FT SIGNAL 1 19
 CC CHAIN 20 291 GAMMA-GLIADIN B.
 CC SEQUENCE 291 AA: 32967 MW: 85732222A2EF041 CRC64;
 SQ
 Query Match 36.6%; Score 332; DB 1; Length 291;
 Best Local Similarity 55.2%; Pred. No. 1.1e-16;
 Matches 69; Conservative 9; Mismatches 23; Indels 24; Gaps 5;
 QY 31 QKPPQPFHPPQPFYLPPELFPQYQIPTPLQPPQPFPPQPPQPPQPPQPPQ 90
 DB 28 QVQWPPQPPQPFPPQPFYLPPELFPQYQIPTPLQPPQPFPPQPPQPPQPPQ 74
 QY 91 QKPPQPFHPPQPFYLPPELFPQYQIPTPLQPPQPFPPQPPQPPQPPQPPQ 146
 DB 75 QKPPQPFHPPQPFYLPPELFPQYQIPTPLQPPQPFPPQPPQPPQPPQPPQ 128
 QY 147 -VPPQ 150
 DB 129 SPFQQ 133
 RESULT 5
 HOG1_HORVU STANDARD; PRT; 105 AA.
 ID HOG1_HORVU
 AC P06472;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE C-hordein (PCP387) (Fragment).
 OS Hordeum vulgare (Barley).
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 CC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae:
 CC Triticeae: Hordeum.
 CC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301876; PubMed=3017812;
 RA Rafalski J.A.;
 RT "Structure of wheat gamma-gliadin genes";
 CC Structure of wheat gamma-gliadin genes";
 CC RT gene 43:221-229(1986).
 CC -I- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -I- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13713; AAA34274.1;
 CC PIR: A25632; EEWG.
 CC InterPro: IPR003612; ANI.
 CC InterPro: IPR001954; G1a_gluTenin.
 CC InterPro: IPR001768; try/amy1_inhbr.
 CC Pfam: PF00234; tryP-alpha_amy1.
 CC PRINTS: PR00208; GLIADGLUTEN.
 CC SMART: SM00499; AAI: 1.
 CC Seed storage protein: Repeat; signal; Multigene family.
 CC FT SIGNAL 1 19
 CC CHAIN 20 291 GAMMA-GLIADIN B.
 CC SEQUENCE 291 AA: 32967 MW: 85732222A2EF041 CRC64;
 SQ

RT common evolutionary origin for two groups of cereal storage protein
 CC genes."
 CC RT EMBL J. 4:9-15(1985).
 CC -I- FUNCTION: SULFUR-POOR SEED STORAGE PROTEIN.
 CC -I- TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X01779; CAA25914.1;
 CC PIR: S07189; S07189.
 CC Seed storage protein; Multigene family.
 CC FT NON_TER 1
 CC SEQUENCE 105 AA: 12180 MW: 3845823443652F CRC64;
 SQ
 Query Match 34.3%; Score 310.5; DB 1; Length 105;
 Best Local Similarity 50.3%; Pred. No. 1.3e-15;
 Matches 73; Conservative 8; Mismatches 23; Indels 41; Gaps 7;
 QY 14 PQQPPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 73
 DB 2 PQQPPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 40
 QY 74 PQQPPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 133
 DB 41 PQQPPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 89
 QY 134 PQQPPLPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 158
 DB 90 LQPHQPPYQQ-----TIMSMV 105
 RESULT 6
 HOG1_HORVU STANDARD; PRT; 305 AA.
 ID HOG1_HORVU
 AC P17990;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Gamma-hordein 1 precursor.
 OS Hordeum vulgare (Barley).
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 CC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae:
 CC Triticeae: Hordeum.
 CC NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301876; PubMed=3017812;
 RA Rafalski J.A.;
 RT "Structure of wheat gamma-gliadin genes";
 CC Structure of wheat gamma-gliadin genes";
 CC RT gene 43:221-229(1986).
 CC -I- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC -I- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
 CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
 CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
 CC NEAR DIRECT REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X01779; CAA25914.1;
 CC PIR: S07189; S07189.
 CC Seed storage protein; Multigene family.
 CC FT NON_TER 1
 CC SEQUENCE 105 AA: 12180 MW: 3845823443652F CRC64;
 SQ

```

CC -----
DR EMBL: X13508; CAA31861.1; -
DR EMBL: M36378; AAA32955.1; -
DR PIR: S08312; S08312.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001954; GLIa-glutenin.
DR InterPro: IPR001768; Try/amyL_inhbt.
DR Pfam: PF00234; try_alpha_amyL; 1.
DR PRINTS: PR00208; GLIADGLUTEN.
DR SMART: SM00499; AAI; 1.
DR Seed storage protein; Signal; Multigene family.
FT SIGNAL 1 19
FT CHAIN 20 305
SQ SEQUENCE 305 AA; 34737 MW; 6D803853FE824AD CRC64;

```

```

Query Match 33.3%; Score 302; DB 1; Length 305;
Best Local Similarity 40.8%; Pred. No. 1,3e-14;
Matches 78; Conservative 13; Mismatches 40; Indels 60; Gaps 9;

```

```

QY 3 QLNPCSGELQSPQSYLQDPYLPQKPFVQGFHTPQGYFYLPEELFPQYQIPPT 62
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 22 QVNP-SVQVQPTQ-----QDPYPS-----QQPFISQSQ-----QQFPQDP 59
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 63 PLQDPQPPQDPQDPPLPDPQPPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 109
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 60 PQDPQPPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 118
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 110 QDPQPPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 147
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 119 QPPQS-YPPPLQPPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 177
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 148 PQQASCIQSMV 158
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 178 -SLSTYMSKI 187
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

RESULT 7

```

ID GDAG_WHEAT STANDARD; PRT; 307 AA.
AC P18573;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Alpha/beta-gliadin M1 precursor (Prolamin).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Chinese Spring; TISSUE=Endosperm;
RA MEDLINE=91346679; PubMed=2102865;
RA Garcia-Maroto F., Manana C., Garcia-Olmedo F., Carbonero P.;
RT "Nucleotide sequence of a cDNA encoding an alpha/beta-type gliadin
RT from hexaploid wheat (Triticum aestivum).";
RL Plant Mol. Biol. 14:867-868(1990).
CC - FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC - MISCELLANEOUS: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
CC HOMOLOGY CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE
CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17361; CAA35238.1; -

```

```

DR PIR: S10015; S10015.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001954; GLIa-glutenin.
DR InterPro: IPR001768; Try/amyL_inhbt.
DR Pfam: PF00234; try_alpha_amyL; 1.
DR PRINTS: PR00208; GLIADGLUTEN.
DR SMART: SM00499; AAI; 1.
DR Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 307
SQ SEQUENCE 307 AA; 35397 MW; 06C1858BD96F1E08 CRC64;

```

```

Query Match 32.0%; Score 289.5; DB 1; Length 307;
Best Local Similarity 31.9%; Pred. No. 9.2e-14;
Matches 81; Conservative 18; Mismatches 40; Indels 115; Gaps 10;

```

```

QY 3 QLNPCSGELQSPQSYLQDPYLPQKPFVQGFHTPQGYFYLPEELFPQYQIPPT 57
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 27 QLPQNPQSQDPQDPVLPVQDQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 83
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 58 YQIPPLQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 110
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 84 PLQPPV-QPQLPVP-QPQLPVPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 138
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 111 QPPQPPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 124
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 139 QQQKQQQQQQQQQLLQQLLPCHDVLQCHSTAYGSSQVLYQDSTYQLVQLCCQQLW 198
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 125 -----PQDPQSYSP-----VQDPQDP 140
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 199 QIPQSRCAITHNVHAIIILHQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 258
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 141 PQ-----QDPVPQ 149
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 259 PQQGSVPQDPQDPQ 272
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

RESULT 8

```

ID HOR8_HORVU STANDARD; PRT; 72 AA.
AC P17991;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE C-hordein (Clone pc HOR1-3) (Fragment).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RA Rasmussen S.K., Brandt A.;
RT "Nucleotide sequences of cDNA clones for C-hordein polypeptides.";
RL Carlsberg Res. Commun. 51:371-379(1986).
CC - FUNCTION: SULFUR-POOR SEED STORAGE PROTEIN.
CC - TISSUE SPECIFICITY: DEVELOPING ENDOSPERM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35610; AAA32943.1; -
DR PIR: A25677; A25677.
DR Seed storage protein; Multigene family.
FT NON_TER 1
SQ SEQUENCE 72 AA; 8590 MW; 79A1B21C6D013B8C CRC64;

```


AC P04722;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha/beta-gliadin A-II precursor (Prolamin).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 NCBI_TaxID=4565;
 RX MEDLINE=85234522; PubMed=2989281;
 RA Okita T.W., Cheesbrough V., Reeves C.D.;
 RT gliadin DNA sequences.";
 RL J. Biol. Chem. 260:8203-8213(1985).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC HOMOLOGUE CLASSES: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M10092; AAA34276.1;
 DR PIR: C22364; C22364.
 DR Interpro: IPR003612; AAI.
 DR Interpro: IPR001954; Glla-glutenin.
 DR Interpro: IPR001768; TRY/amyL_inhbr.
 DR Pfam: PF00234; TRY_alpha_amyL; 1.
 DR PRINTS: PR00208; GLIADGLUTEN.
 DR SMART: SM00499; AAI; 1.
 KM Seed storage protein; Repeat; Signal; Multigene family.
 FT SIGNAL 1 20
 FT CHAIN 21 291 ALPHA/BETA-GLIADIN A-II.
 SO SEQUENCE 291 AA: 33661 MW: 98399380825A2E0 CRC64;
 Query Match 27.6%; Score 250; DB 1; Length 291;
 Best Local Similarity 30.5%; Pred. No. 4.5e-11;
 Matches 72; Conservative 15; Mismatches 45; Indels 104; Gaps 8;
 QY 10 ELDSPOOSYLDQRP-----QNPYLQKPRPVQOFRHTPOQYFPLPELEFPQVQ 59
 DB 29 QLNPSQOQRPQEVYLMQOQOQFPGQOQOFRPQOQRPQOQRPFSQOQ-----PYLDLPFPQVQ 85
 QY 60 ITPRLDQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQ 107
 DB 86 LPVP-----QOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQ 140
 QY 108 ----- 107
 DB 141 IPCRDVYLQOQNTAHGSSQVLDQSTYOLVQOLCCQOQMLQPEQSCQAHHNVHAILHQQ 200
 QY 108 ---FPQOQOQRPQOQRPQOQOQITPOQOQOQSP-----VQOQOQFPQO-----QOQVYVQ 149
 DB 201 QHHHQQOQ 256
 RESULT 14
 GDA3_WHEAT STANDARD: PRT: 282 AA.
 AC P04723;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Alpha/beta-gliadin A-II precursor (Prolamin).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 NCBI_TaxID=4565;
 RX MEDLINE=85234522; PubMed=2989281;
 RA Okita T.W., Cheesbrough V., Reeves C.D.;
 RT gliadin DNA sequences.";
 RL J. Biol. Chem. 260:8203-8213(1985).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC HOMOLOGUE CLASSES: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M1076; AAA34283.1;
 DR PIR: E22364; E22364.
 DR Interpro: IPR003612; AAI.
 DR Interpro: IPR001954; Glla-glutenin.
 DR Interpro: IPR001768; TRY/amyL_inhbr.
 DR Pfam: PF00234; TRY_alpha_amyL; 1.
 DR PRINTS: PR00208; GLIADGLUTEN.
 DR SMART: SM00499; AAI; 1.
 KM Seed storage protein; Repeat; Signal; Multigene family.
 FT SIGNAL 1 20
 FT CHAIN 21 282 ALPHA/BETA-GLIADIN A-II.
 SO SEQUENCE 282 AA: 32236 MW: 87277047784E870A CRC64;
 Query Match 26.7%; Score 242; DB 1; Length 282;
 Best Local Similarity 39.7%; Pred. No. 1.5e-10;
 Matches 69; Conservative 16; Mismatches 49; Indels 40; Gaps 10;
 QY 3 QLNPSQOELDSPOOS---YLQRP---QNPYLQKPRPVQOFRHTPOQYFPLPELEFP 56
 DB 27 QLNQNPQOQ 77
 QY 57 QYQITPRLDQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQ 116
 DB 78 Q---PQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQRPQOQ 129
 QY 117 PQOQOQ-----IIFQOQ-----QOQSYVQOQRPQOQRPQOQRPQOQASC 153
 DB 130 QOILDQOQLPCRDVYLQOQNTAHGSSQVLDQSSYVQOQOQOQOQOQOQOQOQOQ 183
 RESULT 15
 HOR3_HORVU STANDARD: PRT: 264 AA.
 AC P06471;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE B3-hordein (Fragment).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Hordeum.
 NCBI_TaxID=4513;
 RX MEDLINE=85234522; PubMed=2989281;
 RA Okita T.W., Cheesbrough V., Reeves C.D.;
 RT gliadin DNA sequences.";
 RL J. Biol. Chem. 260:8203-8213(1985).
 CC -1- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
 CC HOMOLOGUE CLASSES: THE ALPHA/BETA-GLIADINS CAN BE DIVIDED INTO 5
 CC TO SINGLE BASE SUBSTITUTIONS AND TO DUPLICATIONS OR DELETIONS
 CC WITHIN OR NEAR DIRECT REPEATS. THERE ARE MORE THAN A 100 COPIES OF
 CC THE GENE FOR ALPHA/BETA-GLIADIN PER HAPLOID GENOME.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M1076; AAA34283.1;
 DR PIR: E22364; E22364.
 DR Interpro: IPR003612; AAI.
 DR Interpro: IPR001954; Glla-glutenin.
 DR Interpro: IPR001768; TRY/amyL_inhbr.
 DR Pfam: PF00234; TRY_alpha_amyL; 1.
 DR PRINTS: PR00208; GLIADGLUTEN.
 DR SMART: SM00499; AAI; 1.
 KM Seed storage protein; Repeat; Signal; Multigene family.
 FT SIGNAL 1 20
 FT CHAIN 21 282 ALPHA/BETA-GLIADIN A-II.
 SO SEQUENCE 282 AA: 32236 MW: 87277047784E870A CRC64;

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:44:16 ; Search time 44 Seconds
(without alignments)
345.210 Million cell updates/sec

Title: US-09-743-533-19
Perfect score: 906
Sequence: 1 MROLNPGSOELQSPQSYLQ.....PPQOPQPVPOQASCMSMV 158

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	830.5	91.7	260	2	seed storage prote
2	752	83.0	310	2	C-hordein precursor
3	580	64.0	347	2	probable hordein C
4	566	62.5	357	2	omega secalin prec
5	562.5	62.1	357	2	omega secalin prec
6	412	45.5	327	2	gamma-gliadin prec
7	384.5	42.4	302	2	gamma-gliadin prec
8	347	38.3	251	2	gamma-gliadin prec
9	332	36.6	291	1	EBMTG
10	310.5	34.3	105	2	C-hordein (clone p
11	302	33.3	305	2	gamma-hordein 1 pr
12	289.5	32.0	307	2	alpha/beta-gliadin
13	279.5	30.8	509	2	egg envelope prote
14	272.5	30.1	72	2	C-hordein (pc hor1
15	271.5	30.0	293	2	hordein B1 - barley
16	268.5	29.6	326	2	alpha/beta-gliadin
17	263.5	29.1	297	2	alpha/beta-gliadin
18	262.5	29.0	290	2	alpha/beta-gliadin
19	260.5	28.8	320	2	gamma/beta-gliadin
20	256.5	28.3	286	2	gamma/beta-gliadin
21	255.5	28.2	319	2	alpha/beta-gliadin
22	255	28.1	296	2	alpha/beta-gliadin
23	253.5	28.0	296	2	glutelin - wheat
24	250	27.6	291	2	alpha/beta-gliadin
25	244.5	27.6	1343	2	cell division prot
26	242.5	26.8	374	2	glutelin low molec
27	242	26.7	282	2	alpha/beta-gliadin
28	240	26.5	282	2	B3-hordein (clone
29	239	26.4	313	2	alpha/beta-gliadin

30	238	26.3	356	2	S01992	glutelin low molec
31	238	26.3	359	2	T06982	glutelin low molec
32	237	26.2	288	2	T06282	alpha-gliadin prec
33	236.5	26.1	319	2	A22364	alpha/beta-gliadin
34	236	26.0	286	2	S07923	alpha/beta-gliadin
35	235.5	26.0	276	2	S57656	glutelin low molec
36	234	25.8	286	1	EBMTA	alpha/beta-gliadin
37	234	25.8	292	2	B22364	alpha/beta-gliadin
38	234	25.8	1342	2	G90750	cell division prot
39	234	25.8	1342	2	E85614	cell division prot
40	230.5	25.4	261	2	S57655	glutelin low molec
41	219	24.2	271	2	T04474	glutelin low molec
42	218.5	24.1	68	2	B25677	C-hordein (pc-919)
43	217	24.0	1329	2	A64828	cell division prot
44	211.5	23.3	307	2	S04325	glutelin low molec
45	208.5	23.0	462	1	LUD07	annexin VII - slim

ALIGNMENTS

RESULT 1
S18350
seed storage protein - barley
C:Species: Hordeum vulgare (barley)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S18350
R:Entwistle, J.; Knudsen, S.; Mueller, M.; Cameron-Mills, V.
Plant Mol. Biol. 17, 1217-1231, 1991
R>Title: Amber codon suppression: The in vivo and in vitro analysis of two C-hordein
A:Reference number: S18350; MUID:92032786; PMID:1932695
A:Accession: S18350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <ENT>
A:Cross-references: EMBL:X60037; NID:919000; PIDN:CAA42642.1; PID:g19001
C:Superfamily: gliadin

Query Match 91.7%; Score 830.5; DB 2; Length 260;
Best local similarity 98.7%; Pred. No. 7.7e-53;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 ROLNPGSOELQSPQSYLQPPQNPYLPRKPPVQOPFHTPQPPYLPBELFPQYQIP 61
DB 21 ROLNPGSOELQSPQSYLQPPQNPYLPRKPPVQOPFHTPQPPYLPBELFPQYQIP 80
QY 62 TPLOPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 121
DB 81 TPLOPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 140
QY 122 QIIFQOPQOSYPVQOPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 150
DB 141 QIIFQOPQOSYPVQOPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOPFPQOP 168

RESULT 2
T06211
C-hordein precursor - barley
C:Species: Hordeum vulgare (barley)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-May-2000
C:Accession: T06211; A03355
R:Salnova, O.V.; Mekhedov, S.L.; Zhelnin, L.G.; Khokhlova, T.A.; Anan'ev, E.V.
Genetika 29, 1070-1079, 1993
A>Title: Nucleotide sequence of the barley C-hordein gene.
A:Reference number: Z15537; MUID:93380629; PMID:8396543
A:Accession: T06211
A:Status: preliminary; translated from GR/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-310 <SAI>
A:Cross-references: EMBL:S66938; NID:g442523; PIDN:AA828161.1; PID:g442524
R:Shewry, P.R.; Autran, J.C.; Nimmo, C.C.; Lew, E.J.L.; Kasarda, D.D.
Nature 286, 520-522, 1980
A:Reference number: A93228


```

JS0402
gamma-gliadin precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Nov-1996
C:Accession: JS0402
R:Sugiyama, T.; Katsuki, A.; Soell, D.
Plant Sci. 44, 205-209, 1986
A:Title: The nucleotide sequence of a wheat gamma-gliadin genomic clone.
A:Reference number: JS0402
A:Accession: JS0402
A:Molecule type: DNA
A:Residues: 1-327 <SUG>
A:Experimental source: cv. Yamhill
C:Comment: Gamma-gliadins belong to a subclass of the storage proteins.
C:Superfamily: gliadin
C:Keywords: storage protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-327/Product:gamma-gliadin #status predicted <MAT>

```

[illegible]

```

RESULT 7
JA0153
gamma-gliadin precursor - wheat
N:Alternate names: seed storage protein
C:Species: Triticum aestivum (common wheat)
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 15-Nov-1996
C:Accession: JA0153
R:Schneets, K., Hedgcock, C.
Plant Sci. 57, 141-150, 1988
A:Title: Nucleotide sequence of a gamma gliadin gene: comparisons with other gamma gliad
A:Reference number: JA0153
A:Accession: JA0153
A:Molecule type: DNA
A:Residues: 1-302 <SCCH>
A:Experimental source: seed
C:Comment: Wheat storage proteins are classified as prolamins (alpha-/beta-, gamma- and
lity, molecular weight, and amino acid content.
C:Superfamily: gliadin
F:1-19/Domain: signal sequence #status predicted <SIG>
F:1-20-302/Product: gamma-gliadin #status predicted <GAG>

```

	Query Match	42.4%	Score 384.5;	DB 2;	Length 302;
	Best Local Similarity	58.0%;	Pred. No. 6.3e-21;		
	Matches 91; Conservative	11;	Mismatches 28;	Indels 27;	Gaps 11;
OY	3 QLNPCSGELQSPQSUYLQAPYPQNYLRLOKPRPVQOQPFHT--PQQYPRYLPEELFPQYQ 59	:: ::	::	::	::
Dd	22 QVDP-SGVQWMPQ-----QQPPQ-----QQQP-CQOPORTIPQPMOTHNHQQQFFPOQ 71	:: ::	::	::	::
OY	60 IPRPLQAPQPPP--QQPQQLPRRQPPRPQDQGF--QRPET--PQDPQRP--QQP 112	:: ::	::	::	::
Dd	72 QTTHAQPDQAQFQTQIQDQAPFPQTPQDQQLPRPQDQAPFPQDQAPFPQDQAPFPQSDQP 131	:: ::	::	::	::

QY 113 QQFFPQQPQQLIFQQPQSSYPVQPQQPFPQDPQVPVQ 149
 ||||| | | | | | | | | |
 Db 132 QQFFPQPPQQ--FPQPQ-----QPQQSFPQQQQPAIQ 161

```

RESULT 8
PS0094
gamma-gliadin precursor (clone pm10) - wheat (fragment)
C:Species: Triticum aestivum (common wheat)
C:Date: 31-Dec-1990 #sequence revision 31-Dec-1990 #text:change 20-Aug-1999
C:Accession: PS0094
R:Schneets, K.; Rafalski, J.A.; Hedgcock, C.; Soell, D.G.
Plant Sci. Lett. 37, 221-225, 1985
A:Title: Heptapeptide repeat structure of a wheat gamma-gliadin.
A:Reference number: ps0094
A:Accession: ps0094
A:Molecule type: mRNA
A:Residues: 1-251 <SCH>
A:Cross-references: GB:M16060; NID:q170735; PIDN:AAA34288.1; PID:q170736
C:Comment: Gliadins are the major storage proteins of wheat grain.
C:Superfamily: gliadin
C:Keywords: storage protein
P:1-19/Domain: signal sequence #status predicted <SIG>
P:20-251/Product: gamma-gliadin (clone pm10) (fragment) #status predicted <MAT>

```

```

RESULT 9
EEMTG
gamma-gliadin B precursor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: A25632
R:Rafalski, J.A.
Gene 43, 221-229, 1986
A:Title: Structure of wheat gamma-gliadin genes.
A:Reference number: A25632; MUID:86301876; PMID:3017812
A:Accession: A25632
A:Molecule type: DNA
A:Residues: 1-291 <RAE>
A:Cross-references: GB:M13713; NID:g170707; PIDN:AAA4274.1; PID:g170708
C:Comment: Gliadin, a glutamine and proline rich protein, is a major storage protein
C:Superfamily: gliadin
C:Keywords: seed; storage protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-291/Product: gamma-gliadin B #status predicted <MAV>

```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2003, 14:41:41 : Search time 31 Seconds
(without alignments)
1050.176 Million cell updates/sec

Title: us-09-743-533-19
Perfect score: 906
Sequence: 1 MKOLNPGSOELQSPQOASTYLQ.....PFPQPPQPVQOASCIWSMV 158

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:.*
2: SP_BACTERIA:.*
3: SP_FUNGI:.*
4: SP_HUMAN:.*
5: SP_INVERTEBRATE:.*
6: SP_MAMMAL:.*
7: SP_MHC:.*
8: SP_ORGANELLE:.*
9: SP_PHAGE:.*
10: SP_PLANT:.*
11: SP_PROTOZOA:.*
12: SP_VIRUS:.*
13: SP_VERTEBRATE:.*
14: SP_UNCLASSIFIED:.*
15: SP_VIRUS:.*
16: SP_BACTERIAP:.*
17: SP_ARCHAEA:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	830.5	91.7	260	10 Q40053	040053 hordeum vul
2	752	83.0	310	10 Q41210	041210 hordeum vul
3	580	64.0	347	10 Q40055	040055 hordeum vul
4	574	63.4	280	10 Q9FHW7	09FHW7 triticeum ae
5	568.5	62.7	357	10 Q04655	004655 secale cere
6	566	62.5	357	10 Q05573	005573 secale cere
7	562.5	62.1	357	10 Q05573	005573 secale cere
8	457	50.4	455	10 Q9FR41	09FR41 secale cere
9	440	48.6	337	10 Q94G98	094G98 triticeum ae
10	440	48.6	337	10 Q94G98	094G98 triticeum ae
11	412	45.5	308	10 Q94G96	094G96 triticeum ae
12	412	45.5	327	10 Q94G91	094G91 triticeum ae
13	410.5	45.3	255	10 Q94G91	094G91 triticeum ae
14	393.5	43.4	256	10 Q94G91	094G91 triticeum ae
15	384.5	42.4	302	10 Q94G93	094G93 triticeum ae
16	383.5	42.3	300	10 Q94G93	094G93 triticeum ae

17	383.5	42.3	300	10 Q9FUA1	09FUA1 aegilops 10
18	378.5	41.8	280	10 Q41602	041602 triticeum tu
19	375	41.4	234	10 Q9FTC3	09FTC3 aegilops sh
20	361	39.8	229	10 Q9FTC4	09FTC4 aegilops se
21	361	39.8	231	10 Q9FTC5	09FTC5 aegilops se
22	357	39.4	285	10 Q94G94	094G94 triticeum ae
23	351	38.7	298	10 Q94G92	094G92 triticeum ae
24	350	38.6	294	10 Q9FUA0	09FUA0 aegilops 10
25	347	38.3	243	10 Q9FTC0	09FTC0 aegilops ta
26	347	38.3	259	10 Q9FS62	09FS62 triticeum ae
27	347	38.3	264	10 Q9FS74	09FS74 triticeum ae
28	347	38.3	267	10 Q9FEG9	09FEG9 triticeum ae
29	347	38.3	267	10 Q9FS60	09FS60 triticeum ae
30	347	38.3	270	10 Q9FS69	09FS69 triticeum ae
31	347	38.3	283	10 Q9FS73	09FS73 triticeum ae
32	347	38.3	295	10 Q9XEW0	09XEW0 triticeum ae
33	347	38.3	298	10 Q9SYX8	09SYX8 triticeum ae
34	343	37.9	279	10 Q41543	041543 triticeum ae
35	341.5	37.7	259	10 Q9FS75	09FS75 triticeum ae
36	338	37.3	243	10 Q9FMB9	09FMB9 aegilops ta
37	337.5	37.3	252	10 Q9FS77	09FS77 triticeum ae
38	335	37.0	192	10 Q9FS58	09FS58 triticeum ae
39	335	37.0	250	10 Q9FS76	09FS76 triticeum ae
40	335	37.0	254	10 Q9FS56	09FS56 triticeum ae
41	334	36.9	244	10 Q9FS61	09FS61 triticeum ae
42	332.5	36.7	248	10 Q9FS57	09FS57 triticeum ae
43	332	36.6	192	10 Q9FS59	09FS59 triticeum ae
44	324	35.8	563	13 Q919M8	0919M8 oncorhynch
45	323.5	35.7	277	10 Q9M415	09M415 triticeum ae

ALIGNMENTS

RESULT 1
ID Q40053 PRELIMINARY: PRT: 260 AA.

AC Q40053;
DT 01-NOV-1996 (TREMUREL. 01, Created)
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE Hor-17 C-hordein.
OS Hordeum vulgare (Barley).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Pooideae:
OC Triticeae: Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BOMI;
RX MEDLINE=92032786; PubMed=1932695;
RA Ehlwistle J., Knudsen S., Mueller M., Cameron-Mills V.,
RT "Amber codon suppression: the in vivo and in vitro analysis of two C-
hordein genes from barley.";
RL Plant Mol. Biol. 17:1217-1231(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BOMI;
RX MEDLINE=94035184; PubMed=8220485;
RA Mueller M., Knudsen S.,
RT "The nitrogen response of a barley C-hordein promoter is controlled by
positive and negative regulation of the GCN4 and endosperm box.";
RL Plant J. 4:343-355(1993).
DR EMBL: X60037; CAA42642.1;
SQ
SEQUENCE 260 AA: CB220BD2309280AE CRC64;

Query Match 91.7%; Score 830.5; DB 10; Length 260;
Best Local Similarity 98.7%; Pred. No. 4.1e-69;
Matches 147; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 2 KOLNPGSOELQSPQOASTYLQPPQNPYLPQKPPPVQOFPHTPPQYPPYLPPELFPYQIP 61
Db 21 KOLNPGSOELQSPQOASTYLQPPQNPYLPQKPPPVQOFPHTPPQYPPYLPPELFPYQIP 80

Gy	62	TPLTPOQPFPPQOQDPLPRPQOPFPWMOPOQGFPOPOEFTIPQODQOFFQOPQOPQ	121
Dd	81	TPLTPOQPFPPQOQDPLPRPQOPFPWMOQOGFPOQOEFTIPQODQOPFPQOCPOFFPQOPQ	140
OY	122	QIIFQOPQOSTPVQPQOPFPQOPPOPVQO	150
Db	141	QIIFQOPQOSTPVQPQOPFP-QOPVPQO	168
RESULT 2			
ID	Q41210	PRELIMINARY;	PRT; 310 AA.
AC	Q41210;		
DT	01-NOV-1996 (TREMBLrel_01, Created)		
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel_19, Last annotation update)		
DE	C-hordein.		
GN	C-HORDEIN.		
OS	Hordium vulgare (Barley).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;		
OC	Triticeae; Hordeum.		
OX	NCBI_TaxID=4513;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93380629; PubMed=8396543;		
RA	Sahnova O.V., Mekhedov S.L., Zheleznin L.G., Khokhlova T.A.,		
RA	Anan'ev E.V.;		
RL	"[nucleotide sequence of the barley C-hordein gene].";		
DR	Genetika 29:1070-1079(1993).		
EMBL	S6638; AAA28161.1;		
SO	SEQUENCE 310 AA; 36508 MW; 7A9621869D50D672 CRC64;		

	Query Match	Similarity	83.0%	Score 752:	DB 10:	Length 310;	
	Best Local	Similarity	58.6%	Pred. No. 8, 1e-62;			
	Matches	143:	Conservative	2:	Mismatches	3:	Indels 96; Gaps 2;
QY	2	RQLNPSOGLSPQSSYLQOPYPQNPNVTLFQKRPVVQQPFRHTPQQXYEVLPEELPYQYQITP	61				
Db	21	RQLNPSOGLSPQSSYLQOPYPQNPNVTLFQGFPPVQOGPHRTFPQQXVFYUPLPEELSPQYQITP	80				
QY	62	TPLDLPQOPPPPQQPQQPLPRPQQPFFWQQPQQPFFRPPQQDEFLPQ-----	102				
Db	81	TPLDLPQOPPPPQQPQQPLPRPQQPFFWQQPQQPPIPVQDPGQPFNQDPQOIISQDPQ	140				
QY	103	-----	102				
Db	141	QPFPQQPQQPFPQQPQQPFWQQPQQPFPQQPQQPFFLQDPQQPFFWQQPQQPFPFPQQPQTIANHQP	200				
QY	103	-----QPQQPFPQQDPQQPFPQQPQQIIFQDFQSSYVPQQPFPQQPQP	146				
Db	201	QQPFSsqQQPQQPFLqQQPQQPFPQQQqQQPFPQQQIIIFQDPQSSYVPQQPQPFp-QPQP	259				
QY	147	VPPQ 150 					
Db	260	VPPQ 263					
RESULT 3							
ID	Q40055	PRELIMINARY:	PRT:	347	AA.		
AC	Q40055;						
Dt	01-NOV-1996	(TREMBlrel. 01, Created)					
Dt	01-NOV-1996	(TREMBlrel. 01, Last sequence update)					
Dt	01-DEC-2001	(TREMBlrel. 19, Last annotation update)					
De	C hordeum precursor.						
OS	Hordeum vulgare (Barley).						
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;						
OX	Triticeae; Hordeum.						
OX	NCBI_TaxID=4513;						
RN	[1]						

RP	SEQUENCE FROM N.A.			
RC	STRAIN-BOMI; TISSUE=IMATURE; ENDOSPERM;			
RX	MEDLINE=89351278; PubMed=3255313;			
RA	Eptwistle J.;			
RT	"Primary structure of a C-hordein gene from barley.";			
RL	Carlberg Res. Commun. 53:247-258(1988).			
DR	EMBL; M36941; AAA92333.1; -.			
KW	Signal.			
FT	1	20	POTENTIAL.	
FT	21	347	C HORDEIN.	
SO	SEQUENCE	347 AA; 40546 MW;	1E48319B;BCBCD9D CRC64;	

[illegible]

RESULT 4	
09FWU7	PRELIMINARY; PRT; 280 AA.
ID	09FWU7
AC	09FWU7;
DT	01-MAR-2001 (TIREMBLrel. 16, Created)
DT	01-MAR-2001 (TIREMBLrel. 16, Last sequence update)
DT	01-DEC-2001 (TIREMBLrel. 19, Last annotation update)
DE	Omega gliadin storage protein.
OS	Triticum aestivum (Wheat).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OX	Triticeae; Triticum.
NCBI_TaxID=4565;	
NP	[1]
RP	SEQUENCE FROM N.A.
RA	Hsli C.C., Anderson O.D.;
RT	"Isolation and characterization of wheat omega gliadin genes.";
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF280605; AAC17702.1; -.	
SO	SEQUENCE 280 AA; 32554 MW; E2FB865C8CE4585D CRC64;

[illegible]

RESULT 5
004365

ID 004365; PRELIMINARY; PRT; 357 AA.

AC 004365;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Omega secalin.

OS Secale cereale (Rye).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Secale.

OC NCBI_taxid=4550;

OX [1]

RN SEQUENCE FROM N.A.

RP TISSUE-ENDOSPERM.

RC MEDLINE=97094352; PubMed=8939819;

RX Clarke B.C., Mukai Y., Appels R.;

RA "The Sec-1 locus on the short arm of chromosome 1R of rye (Secale cereale).";

RT RT

RL Chromosome 105:269-275(1996).

RN [2]

RN SEQUENCE FROM N.A.

RP TISSUE-ENDOSPERM;

RC Clarke B., Appels R.;

RA "Sequence variation at the Sec-1 locus of rye.";

RT Plant Syst. Evol. 0:0-0(1997).

DR EMBL: AF000227; AAB58403.1;-

SO SEQUENCE 357 AA; 41625 MW; E3219653C105EFP6 CRC64;

	Query Match	62.7%	Score 568.5	DB: 10	Length 357	
	Best Local Similarity	68.6%	Pred. No. 7.1e-45			
	Matches 109;	Conservative 11;	Mismatches 30;	Indels 9;	Gaps 3;	
OY	2 ROLNFCSEGLSPQSSSYLQRPYPNVLPQRKRPVQGRFHFRRPOXYFPYLPEELFRQYQLP 61					
Db	20 ROLNFBSEGLSP-----QQPVKPKQSYPQDPYPFHSGFFPRFOGXSPVQDPPFPQDP 74					
OY	62 TPLOPQGFPPQDPQPLRPPQDPFRWQDQGRPPROGEPFRPQDPQPFQDPQDP---FPQ 118					
Db	75 TPLOPQGFPPQDPQDPFSQPDQDLPLQDQGRFPQDPQDPQDPQDPQDPQDPQDPQDPQDP 134					
OY	119 QPQQLIFQDPQDSYVPVQDPQDPFPQDPQ--PVPQASCIWS 156					
Db	135 QPQQLIFQDTQDPFPLQDPQDPFPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDPQDP 173					
RESULT 6						
ID	005573	PRELIMINARY:	PRT:	357 AA.		
AC	005573;					
DT	01-NOV-1996 (TREMBlrel. 01, Created)					
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)					
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)					
DE	Sec1 precursor.					
GN	Sec1					
OS	Secale cereale (Rye).					
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;					
OC	Triticeae; Secale.					
OX	NCBI_TaxID=4550;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=CV_CAZELLE; TISSUE=ENDOSPERM;					
RX	MEDLINE=92032773; PubMed=1932685;					
RA	Hull G.A., Hallford N.G., Kreis M., Shewry P.R.:					
RT	"Isolation and characterisation of genes encoding rye prolamins					
RL	containing a highly repetitive sequence motif.";					
RT	Plant Mol. Biol. 17:1111-1115(1991).					
DR	EMBL: X60295; CAA42837.1; -					
KM	Signal.					
FT	SIGNAL	1	19	POTENTIAL.		
FT	CHAIN	20	357	POTENTIAL.		
SO	SEQUENCE	357 AA;	41439 MW;	8EC406DA8F72909E CRC64;		

Query Match	Score	DB	Length
Best Local Similarity	70.7%	Pred. No. 1,2e-44;	
Matches	104;	Conservative	11; Mismatches 24; Indels 8; Gaps 2
QY	2	ROLNPNCSOELQSPQOOSYLDQPYPNQVYLQKQKFFPVQVQOPFHHTQOQYFFPYLPEELFPQYQIP	61
DB	20	ROLNPNSEQLQSP-----QQPYKPKBQSYQOQYPSHQPPFPFOQYSFPYQOQFPFOQPP	74
QY	62	TPQLQOQPPFPQOQOPLPPQOQPFWMQOQPPFPQOQEPILPQOQOQFPFOQPOOP--FPQ	118
DB	75	APLPQOQPPFPQOQOQPFQOQOPLPQOQOQPFQOQPLQOQOQSFQOQOQPPQOQPPQOQPPQOQ	134
QY	119	OPQOQIFQOQOOSYQVQOQOQPFQOQO	145
DB	135	OPQOQIFQOQOQOQOQPFQOQOQPFQOQO	161

RESULT 7			
ID	043639	PRELIMINARY:	PRT: 357 AA.
AC	043639;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Sec1 precursor.		
GN	Sec1.		
OS	Secale cereale (Rye).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;		
OC	Triticeae; Secale.		
OX	NCBI_Taxid=4550;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=CV. GAZELLE; TISSUE=ENDOSPERM;		
RX	MEDLINE=92032773; PubMed=1932685;		
RA	Hull G.A., Halford N.G., Kreis M., Shewry P.R.;		
RT	"Isolation and characterisation of genes encoding rye prolamins		
RT	containing a highly repetitive sequence motif.";		
RL	Plant Mol. Biol. 17:111-115(1991).		
DR	EMBL: X60294; CAA42836.1; -;		
KW	Signal.		
FT	SIGNAL		
FT	CHAIN	1	19
FT		20	357
FT			POTENTIAL.
SO	SEQUENCE	357 AA:	41541 MW; 795705CA10D5B1E CRC64;

	Query Match	62.1%	Score 562.5;	DB 10;	Length 357;	
	Best Local Similarity	67.9%;	Pred. No. 2;Se-44;			
	Matches 108;	Conservative 11;	Mismatches 31;	Indels 9;	Gaps 3;	
Oy	2 ROLNCSOELOSPOQSXYLOQRYRONMPLKRRFPQQQRPNTRPOOXYFRPLYREELFPROXDT 61		: :			
Dd	20 ROLNCSOELOSPO-----OOVPKROGSTRPOORPSHQRPRTPQOYSPYOQOPRFPQOPR 74					
Oy	62 TPLDQQRPFPPQOPRLPFRPQRPFRMPQOPRFOERPTPOQROPQOPFPQOPPOOP---FPQ 118	T :				
Dd	75 TPLOQOPRRPQRPQOPRFPQOPROODLPLDRPOQSFQONRNLPQOPQOSFFQOPRQOPRQOPRO 134	T P L O Q O P R R P Q R P Q O P R F P Q O P R O O D L P L D R P O Q S F Q O N R N L P Q O P Q O S F F Q O P R Q O P R Q O P R O				
Oy	119 QPQIIFQOPQGSYPVQRPQOPRFPQOPQ--PVQPOASCIWS 156					
Dd	135 QPQIILPQOTQOPFLQRPQOPRFPQOPRQOPRFAQDPRIIS 173					
 RESULT 8 O9FR41 PRELIMINARY: PRT: 455 AA.						
ID	O9FR41					
NC	O9FR41.					
DT	01-MAR-2001 (TREMBLrel. 16, Created)					
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)					
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)					
DE	Secalin.					
GN	GSEC2-1.					
OC	Secale cereale (Rye).					
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Secale.
 OX NCBI_TaxID=4550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. IMPERIAL;
 RA Murray F.R., Skerrett J.H., Appels R.;
 RT "A gene from the Sec2 (Gli-R2) locus of a wheat 2RS.2BL chromosomal
 translocation line."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF201084; AAG35598.1; -
 DR InterPro: IPR001954; G1a.gluTenin.
 DR InterPro: IPR002965; P.rich.extensin.
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.
 DR PRINTS: PR00208; GLTADGLUTEN.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 FT CHAIN 20 455
 SQ SEQUENCE 455 AA; 51731 MW; C1A176DD045D927 CRC64;

Query Match 50.4%; Score 457; DB 10; Length 455;
 Best Local Similarity 61.8%; Pred. No. 1,6e-34;
 Matches 94; Conservative 12; Mismatches 34; Indels 12; Gaps 6;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOPFHTPQOYPVYLPBELFPOYOIPT 62
 Db 22 QVNP-SGVQCCQOQPPFPQOQSSFPQOQPPQOS-----QQFPFPQOQSSFPQOP- 74
 OY 63 PLQOPFPQOQOPLP-RPQOPFPQOQOPLP-QPQEPQOQOQPPFPQOQOQPPFPQOP 120
 Db 75 -YPQOPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOP 132
 OY 121 QQIIFQOQOQSPVQOQOQPPFPQOQ-QPVPQQA 151
 Db 133 QQFPQOQPLQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQ 164

RESULT 9

ID 094G98 PRELIMINARY; PRT: 241 AA.
 AC 094G98;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Gamma-gliadin (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D., Hsia C., Torres V.;
 RT "The wheat gamma-gliadin genes: characterization of ten new sequences
 and further understanding of gamma-gliadin gene family structure."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF234642; AAK84772.1; -
 FT NON_TER 241
 SQ SEQUENCE 241 AA; 28079 MW; A2765A48A35FD6D3 CRC64;

Query Match 48.6%; Score 440; DB 10; Length 241;
 Best Local Similarity 57.5%; Pred. No. 3.2e-33;
 Matches 100; Conservative 12; Mismatches 26; Indels 36; Gaps 11;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOP---FHTPQOYFPLYPEELF----- 55
 Db 22 QVNP-SGVQOQWQO---QQFPQO---PQOPFS-QQPOQIFPQOQOTPHQOQAFQOPQ 72
 OY 56 -----PQOYITPPLQOQOQPPFPQOQOPLP---RPQOPFPQOQOQPPFPQOPQOQ-PIP 101
 Db 73 QTFPHQOQOQFPQOQOQPPFPQOQOQFPQOQOQFPQOQOQFPQOQOQFPQOQOQFPQOP 132

OY 102 QQOPQFPQOQOQPF---PQOPQOII---FQOPQO---SYVPQOQOQPPFPQOQOQPPVQ 149
 Db 133 QQPOLFPQOQOQFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFIQ 186

RESULT 10

ID 094G96 PRELIMINARY; PRT: 337 AA.
 AC 094G96;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Gamma-gliadin.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RA Anderson O.D., Hsia C., Torres V.;
 RT "The wheat gamma-gliadin genes: characterization of ten new sequences
 and further understanding of gamma-gliadin gene family structure."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF234644; AAK84774.1; -
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.
 SQ SEQUENCE 337 AA; 38634 MW; 524DCFA9CE1F0C13 CRC64;

Query Match 48.6%; Score 440; DB 10; Length 337;
 Best Local Similarity 57.5%; Pred. No. 4.4e-33;
 Matches 100; Conservative 12; Mismatches 26; Indels 36; Gaps 11;

OY 3 QLNPCSOELQSPQSYLQDPYQNPYLPQKPPVQOP---FHTPQOYFPLYPEELF----- 55
 Db 22 QVNP-SGVQOQWQO---QQFPQO---PQOPFS-QQPOQIFPQOQOTPHQOQAFQOPQ 72
 OY 56 -----PQOYITPPLQOQOQPPFPQOQOPLP---RPQOPFPQOQOQPPFPQOPQOQ-PIP 101
 Db 73 QTFPHQOQOQFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOP 132
 OY 102 QQOPQFPQOQOQPF---PQOPQOII---FQOPQO---SYVPQOQOQPPFPQOQOQPPVQ 149
 Db 133 QQPOLFPQOQOQFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFPQOQOQPPFIQ 186

RESULT 11

ID 09M6P7 PRELIMINARY; PRT: 308 AA.
 AC 09M6P7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Gamma-gliadin (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YAMHILL;
 RA Arentz-Hansen E.H., McAdams S.N., Molberg O., Kristiansen C.,
 RA Sollid L.M.;
 RT "Production of a panel of recombinant gliadins for the
 RT characterization of T cell reactivity in coeliac disease."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF175312; AAF42989.1; -
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001954; G1a.gluTenin.
 DR InterPro: IPR001768; Try/amyL_inhtr.
 DR Pfam: PF00234; tryp_alpha_amyL.1.

DR PRINTS: PR00208; GLIADGLUTEN.

DR SMART: SM00499; AA1; 1.

FT NON_TER 1

SO SEQUENCE 308 AA; 35198 MW; 0A772815C23BA203 CRC64;
Query Match 45.5%; Score 412; DB 10; Length 308;
Best Local Similarity 50.0%; Pred. No. 1.5e-30;
Matches 97; Conservative 12; Mismatches 37; Indels 48; Gaps 10;

OY 3 OLNPSCOE--LQSPQSYLQOPYPQNPYLQKPFPPQGFHPPQOYFPLPELFPQYQI 60
DB 3 QVDPSSQVGMLOQQLVPLQQLPSQP-----QQTFPPQQTFFPHQPPQOVPQPPQ 53
OY 61 PTP--LQPPQPPFPQPPQPLP---RPQPPFWQPPQPPF---QPQEP1PQPPQPPF--Q 110
DB 54 PQQPF1QPPQPPFPQPPQPPFPQPPQPPQPPQPPFPQPPQPPFPQPPQPPFPQPPQ 113
OY 111 QPQPPFP--QPPQQLIFQPPQOS--YPPVQPPFPQPPQPPVPPQA----- 151
DB 114 QPQPPFPQ1QPPQPPQ--PFPQPPQQLPQPPQPPQSPQPPQPP1QPS1QQLNPCKNILLQ 172
OY 152 -----SC1MSMV 158
DB 173 CKPASLVSLSMSII 186

RESULT 12

O94G91 PRELIMINARY: PRT; 327 AA.

AC Q94G91;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Gamma-gliadin.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
RN NCBI_TaxID=4565;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RA Anderson O.D., Hala C., Torres V.;
RT "The wheat gamma-gliadin genes: characterization of ten new sequences
and further understanding of gamma-gliadin gene family structure."
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF234650; AAK84779.1; -;
DR InterPro: IPR001768; Try/amy1_inhtr.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
SO SEQUENCE 327 AA; 37169 MW; E46B6A4CB5BDB8C2 CRC64;

Query Match 45.5%; Score 412; DB 10; Length 327;
Best Local Similarity 50.0%; Pred. No. 1.6e-30;
Matches 97; Conservative 12; Mismatches 37; Indels 48; Gaps 10;

OY 3 OLNPSCOE--LQSPQSYLQOPYPQNPYLQKPFPPVQPPFHTPPQYFPLPELFPQYQI 60
DB 22 QVDPSSQVGMLOQQLVPLQQLPSQP-----QQTFPPQQTFFPHQPPQOVPQPPQ 72
OY 61 PTP--LQPPQPPFPQPPQPLP---RPQPPFWQPPQPPF---QPQEP1PQPPQPPF--Q 110
DB 73 PQQPF1QPPQPPFPQPPQPPFPQPPQPPQPPQPPFPQPPQPPFPQPPQPPFPQPPQ 132
OY 111 QPQPPFP--QPPQQLIFQPPQOS--YPPVQPPFPQPPQPPVPPQA----- 151
DB 133 QPQPPFPQ1QPPQPPQ--PFPQPPQQLPQPPQPPQSPQPPQPP1QPS1QQLNPCKNILLQ 191
OY 152 -----SC1MSMV 158
DB 192 CKPASLVSLSMSII 205

RESULT 13

O9FTC2
ID O9FTC2 PRELIMINARY: PRT; 255 AA.

AC O9FTC2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Gamma-gliadin (Fragment).
GN GAG56.
OS Aegilops speltoides (Goat grass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Aegilops.
RN NCBI_TaxID=4573;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ACCESSION PI 449340;
RA von Bueren M.;
RT "Polymorphisms in two homeologous gamma-gliadin genes and the
evolution of cultivated wheat."
RL Genet. Resour. Crop Evol. 0:0-0(0).
DR EMBL: AJ389719; CAC10623.1; -;
DR InterPro: IPR001954; G1a-glutenin.
DR InterPro: IPR001768; Try/amy1_inhtr.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00208; GLIADGLUTEN.
FT NON_TER 1
FT NON_TER 255
SO SEQUENCE 255 AA; 30265 MW; E05A13999DE893DF CRC64;

Query Match 45.3%; Score 410.5; DB 10; Length 255;
Best Local Similarity 61.7%; Pred. No. 1.7e-30;
Matches 87; Conservative 8; Mismatches 25; Indels 21; Gaps 7;

OY 21 QPYPNPYLQKPFPPVQPPFHTPPQYFPLPELFPQYQ1PTP--LQPPQPPFPQPPQL 78
DB 2 QPFSQP-----QQTYPPQQTFFPHQPPQPPQPPQOQFLQPPQPPFPQPPQPPY 52
OY 79 P-RPQPPFP--WPQPPFP---QPQEP1PQPPQPPFPQPPQPPF---QPQQLIF--FPQ 128
DB 53 PQQPPFPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 112
OY 129 QQSPVQPPQPPFPQPPQPPVPPQ 149
DB 113 QQPPQPPQPPFPQPPQPPQPPQ 133

RESULT 14

O9FTC1 PRELIMINARY: PRT; 256 AA.

AC O9FTC1;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Gamma-gliadin (Fragment).
GN GAG56.
OS Aegilops speltoides (Goat grass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Aegilops.
RN NCBI_TaxID=4573;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ACCESSION PI 542245;
RA von Bueren M.;
RT "Polymorphisms in two homeologous gamma-gliadin genes and the
evolution of cultivated wheat."
RL Genet. Resour. Crop Evol. 0:0-0(0).
DR EMBL: AJ389720; CAC10624.1; -;
DR InterPro: IPR001954; G1a-glutenin.
DR InterPro: IPR001768; Try/amy1_inhtr.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00208; GLIADGLUTEN.
FT NON_TER 1

FT NON_TER 256 256
 SO SEQUENCE 256 AA; 30456 MM; 8D2B52ED7B6CF9E0 CRC64;

Query Match 43.4%; Score 393.5; DB 10; Length 256;
 Best Local Similarity 51.4%; Pred. No. 6,4e-29;

Matches 94; Conservative 13; Mismatches 21; Indels 55; Gaps 13;

```
OY 12 QSPQSYL--QQYPPQNPYLKQKFFPYQQGFPHPPQGYFYL--PEELFPQYQIPTPLPQ 67
   | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 32 QQPQOQFLQPQPPPPQ--PQPPYPQO-----PQPPPTQPPQLPQSO-----QPQ 78
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 68 QPFPQPPQPPPLP--RPQPPFP--WQPQPPF--QPPPIPPQPPQPPFP--QPPQPPPPQ 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 79 QPYPPQPPQPPYPPQPPQPPFPPTQPPQLFPQSQPPQPPYPPQPPQPPFPQPPQPPQPPFPQ- 137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 120 PQQIIFPQPPQSY--PVPQPPFPQPP--QVPYQO-----ASCIW 155
   : | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 -----YQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 156 SMV 158
   | | |
Db 193 SMT 195
```

RESULT 15

```
O94G93 PRELIMINARY: PRT: 302 AA.
AC O94G93:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Gamma-gliadin.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHEYENNE;
RA Anderson O.D., Hsia C., Torres V.;
RT "The wheat gamma-gliadin genes: characterization of ten new sequences
RT and further understanding of gamma-gliadin gene family structure.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF234647; AAK84777.1; -.
DR InterPro: IPR001768; Try/amyL_inhlt.
DR Pfam: PF00234; try_alpha_amiL.1
SQ SEQUENCE 302 AA; 34286 MM; 663B37C768BAE782 CRC64;
```

Query Match 42.4%; Score 384.5; DB 10; Length 302;

Best Local Similarity 58.0%; Pred. No. 5.1e-28;

Matches 91; Conservative 11; Mismatches 28; Indels 27; Gaps 11;

```
OY 3 QLNPCSQEIQSPQSYLQPPYPPQPPYLPQKPPPVQOQPHHT---PQYFPYLPBELFPQYQ 59
   | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 22 QYDP--SGQVQWPQ---QQPPQ---PQPP--COQPPQTIPQPHQTFHMQPQTFPPQ 71
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 60 IPTPLPQPPFP--QQPPQLPFPQPPFPWQPPQPPF--OPEPI--PQPPQPPFP--QPP 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 72 QTYPHQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 131
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 113 QQPPPPQPPQIIFQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 149
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 QQPPFPQPPQO--FPQPPQ-----QPQSPFPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 161
```

Search completed: May 27, 2003, 14:46:48
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:47:50 ; Search time 518.477 Seconds

(without alignments)
7858.394 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140

Sequence: 1 atagatacagcagatgcctccc.....gagctcgattcgccctata 140

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Geneml:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_in:*

32: em_htg_lnv:*

33: em_htg_other:*

34: em_htg_mus:*

35: em_htg_pin:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_hgo_hum:*

40: em_hgo_mus:*

41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.6	25.4	388	8	HVE1393VR
2	35.6	25.4	558	8	HVE1326PR
3	35.2	25.1	1222	6	AR170979
4	34.4	24.6	2050	1	AP180731
5	34.2	24.4	163	6	AR210350
6	33.8	24.1	484	6	AR147996
7	33.8	24.1	484	6	AR147996
8	33.4	23.9	1271	8	GCN132219
9	33.4	23.9	1447	8	GCN132212
10	33.4	23.9	5506	6	AR195508
11	33.4	23.9	5507	8	LSRGC2D1
12	33.3	23.6	345	9	HSZ96972
13	32.6	23.3	163	6	AR210351
14	32.6	23.3	15394	2	AC107595
15	32.4	23.1	278	8	HV1371VRC
16	32.4	23.1	500	11	G74883
17	32.4	23.1	170241	2	AC016434
18	32.2	23.0	481	9	HSPPT06
19	32.2	23.0	700	6	AX392015
20	32.2	23.0	4350	12	AP173954
21	32.2	23.0	241901	2	AC128708
22	32.2	22.9	541	9	CJSGP2
23	32.2	22.9	672	11	G68272
24	32.2	22.9	1432	3	HSC299386
25	32.2	22.9	3027	8	AP395756
26	32.2	22.9	3551	12	AR122058
27	32.2	22.9	15317	9	AC016995
28	32.2	22.9	1843	10	AP232061
29	31.8	22.4	490	8	CSA416715
30	31.4	22.4	608	1	UBA011367
31	31.4	22.4	7823	12	AF041426
32	31.4	22.4	8696	12	AF379854
33	31.4	22.4	40544	1	SCH5
34	31.4	22.4	1611	8	AF308658
35	31.2	22.3	215	6	AX156393
36	31.2	22.1	419	6	AR166312
37	31.2	22.1	420	6	AR166319
38	31.2	22.1	502	3	U4L27427
39	31.2	22.1	530	6	AX027636
40	31.2	22.1	597	6	AR148006
41	31.2	22.1	597	6	AX116848
42	31.2	22.1	624	8	OS040814
43	31.2	22.1	632	8	BNA489198
44	31.2	22.1	741	8	BNA489197
45	31.2	22.1	741	8	BNA489197

ALIGNMENTS

RESULT 1

LOCUS HVE1393VR 388 bp DNA linear PLN 07-JUL-2002

DEFINITION H. vulgare BARE-1 long terminal repeat DNA; ID E1333VR.

ACCESSION Z84569.1

VERSION Z84569.1 GI:2598547

KEYWORDS BARE-1 retrotransposon; long terminal repeat.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

REFERENCE 1 (bases 1 to 388)

AUTHORS Suontemi, A., Schmidt, D., and Schulman, A.H.

TITLE BARE-1 insertion site preferences and evolutionary conservation of

RNA and cDNA processing sites
Genetica 100 (1-3), 219-230 (1997)

JOURNAL
MEDLINE
98103449
9440275
PUBMED
2 (bases 1 to 388)

REFERENCE
AUTHORS
TITLE
JOURNAL
Schulman A.H.
Direct Submission
Submitted (24-JAN-1997) Schulman A.H., University of Helsinki,
Institute of Biotechnology, P.O. Box 56, Biocentre 1, Viikinkaari
9, FIN-00014 Helsinki, Finland
Location/Qualifiers

FEATURES
source
1..388
/organism="Hordium vulgare"
/db_xref="taxon:4513"

LTR
142..>273
/note="matches nt 1-131 of BARE-1 5' LTR, 5' terminus"

repeat_region
142..273
/transposon="BARE-1"

BASE COUNT
102 a 94 c 87 g 105 t

ORIGIN

Query Match
Best Local Similarity 82.0%; Score 35.6; DB 8; Length 388;
Pred. No. 0.1;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY
11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAAACC 60
|||||
22 GCATGCTCCCGCGCCATGCGCGGATGTCGCTCGCCCACTACATC 71

RESULT 2
LOCUS
HVE1326PR 558 bp DNA linear PLN 07-JUL-2002
H.vulgare BARE-1 long terminal repeat DNA; ID E1326PR.
284562
284562.1 GI:2598543
BARE-1 retrotransposon: long terminal repeat.
Hordium vulgare.
Hordium vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Focaceae;
Poideae; Triliceae; Hordium.
1 (bases 1 to 558)
Suomela A., Schmidt D. and Schulman A.H.
BARE-1 insertion site preferences and evolutionary conservation of
RNA and cDNA processing sites
Genetica 100 (1-3), 219-230 (1997)

JOURNAL
MEDLINE
98103449
9440275
PUBMED
2 (bases 1 to 558)

REFERENCE
AUTHORS
TITLE
JOURNAL
Schulman A.H.
Direct Submission
Submitted (24-JAN-1997) Schulman A.H., University of Helsinki,
Institute of Biotechnology, P.O. Box 56, Biocentre 1, Viikinkaari
9, FIN-00014 Helsinki, Finland
Location/Qualifiers

FEATURES
source
1..558
/organism="Hordium vulgare"
/db_xref="taxon:4513"

LTR
<135..300
/note="matches nt 1682-1829 of BARE-1 5' LTR, 3' terminus"

repeat_region
<153..300
/transposon="BARE-1"

misc_feature
417..548
/note="flanking genomic sequence matches BARE-1a
retrotransposon (Z17327) nt 308-439"

BASE COUNT
139 a 133 c 129 g 157 t

ORIGIN

Query Match
Best Local Similarity 82.0%; Score 35.6; DB 8; Length 558;
Pred. No. 0.11;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY
11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGCACTAAACC 60
|||||
110 GCATGCTCCCGCGCCATGCGCGGATGTCGCTCGCCCACTACATC 159

Db

RESULT 3
LOCUS
AR170979/c 1222 bp DNA linear PAT 17-DEC-2001
AR170979
Sequence 83 from patent US 6297007.
DEFINITION
AR170979
AR170979
VERSION
AR170979.1 GI:17909929

KEYWORDS
SOURCE
ORGANISM
Unknown.
Unknown.
Unclassified.

REFERENCE
1 (bases 1 to 1222)
Waters B., Miao Y., Ho Y., Mai, and Tong S. Kah.
Method for isolation of biosynthesis genes for bioactive molecules
Patent: US 6297007-A 83 02-OCT-2001;
Location/Qualifiers

FEATURES
source
1..1222
/organism="unknown"

BASE COUNT
272 a 361 c 319 g 270 t

ORIGIN

Query Match
Best Local Similarity 92.5%; Score 35.2; DB 6; Length 1222;
Pred. No. 0.17;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY
11 GCATGCTCCCGCGCCATGCGCGGATGTCATGAGG 50
|||||
1176 GCATGCTCCCGCGCCATGCGCGGATGTCGAGCAGC 1137

Db

RESULT 4
LOCUS
AF180731 2050 bp DNA linear BCT 01-OCT-1999
AF180731
Klebsiella pneumoniae dihydrofolate reductase (dhfrXII) and
streptomycin 3'-adenylyltransferase (aadA2) genes, complete cds;
and unknown gene.
DEFINITION
AF180731
AF180731.1 GI:6003514
Klebsiella pneumoniae.
Klebsiella pneumoniae.
Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
Klebsiella.
1 (bases 1 to 2050)

REFERENCE
AUTHORS
TITLE
JOURNAL
Kim J.
Direct Submission
Submitted (26-AUG-1999) Microbiology, Coll. Med., Dankook
University, San29, Anseo-dong, Cheonan, Choongnam 330-180, Republic
of Korea
Location/Qualifiers

FEATURES
source
1..2050
/organism="Klebsiella pneumoniae"
/strain="Kps15"
/db_xref="taxon:573"
/transposon="class I integron"

misc_feature
128..131
137..634
/note="putative integrase target site; GTTA-sequence"

gene
137..634
/gene="dhfrXII"

CDS
137..634
/gene="dhfrXII"
/function="trimethoprim resistance"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AA00479.1"
/db_xref="GI:6003515"
/translation="MNSSESVRIYLVAAAGANKRVITGNGFNIPWKIPGEOKIFRRITLGEK
VVMGRKTFESIGKPLPNRHVLIISROANYRATGCVVSTLSHAIALASDELNYVA

```

GAAGTATLALPHANGVFLSEVHQTFEGDAFFPMLNETEFELVSTETIQAVIPYTHSV
ARNRG"
stem_loop      648..679
misc_feature    712..715
                /note="putative integrase target site; GTTA-sequence"
CDS             746..1036
                /note="orf"
                /codon_start=1
                /transl_table=1
                /product="unknown"
                /protein_id="AAF00480.1"
                /db_xref="GI:6003516"
                /translation="METQAFSFGVIOCLFCLPSGLHGLRRFSVFLASSPCVASA
                1032..1035
                /note="putative integrase target site; GTTA-sequence"
                1044..1833
                /gene="aada2"
                1044..1047
                /gene="aada2"
                1054..1833
                /gene="aada2"
                /note="aminoacyl-adenyl transferase A"
                /codon_start=1
                /transl_table=1
                /product="streptomycin 3'-adenyl transferase"
                /protein_id="AAF00481.1"
                /db_xref="GI:6003517"
                /translation="MTLEVSQSLSEVSVIERHLESTLLAVHLGSAVDOGKPYSDI
                DLIVYAVKLDDETRRALNDMEASAFPGSEETLRAIEVLVYHDDITPWRPAKRE
                LQFGEWQNDILAGIPEPAMIDILAILLKARHSVALVGPAAEEFEDVPEDDELE
                ALRPLTKLMSQVPMAGDERNVVLTLSINVSATGKIAEPDVADCAIKRLPAOYOP
                VLLEAKQAVYLQCKDHLASRADHLEEFIRFVKETIKSVSK"
                1836..1891
stem_loop      468 a 545 c 525 g 512 t
BASE COUNT
ORIGIN
Query Match      24.6%; Score 34.4; DB 1; Length 2050;
Best Local Similarity 97.2%; Pred. No. 0.38;
Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCAT 46
    1 GCATGCTCCGGCGCCATGCGCGGATTCAT 36
Db 1 GCATGCTCCGGCGCCATGCGCGGATTCAT 36

RESULT 5
LOCUS AR210350 163 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 51 from patent US 6387657.
ACCESSION AR210350
VERSION AR210350.1 GI:21512558
KEYWORDS
ORGANISM Unknown.
SOURCE Unknown.
REFERENCE Unclassified.
1 (bases 1 to 163)
AUTHORS Bocstein,D.A., Cohen,R.L., Goddard,A.D., Gurney,A.L., Hillan,K.J.,
TITLE Lawrence,D.A., Levine,A.J., Pennica,D., Roy,M.Am. and Wood,W.I.
JOURNAL WISP polypeptides and nucleic acids encoding same
FEATURES
    source      1..163
                Location/Qualifiers
BASE COUNT 35 a 45 c 46 g 37 t
ORIGIN
Query Match      24.4%; Score 34.2; DB 6; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.27;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAG 49
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAG 49
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAG 49

```

```

Db 41 GCATGCTCCGGCGCCATGCGCGGATTCATCACTAG 79

RESULT 6
LOCUS AR147996 484 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 175 from patent US 6225054.
ACCESSION AR147996
VERSION AR147996.1 GI:15112086
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 484)
AUTHORS Frudakis,T.N., Smith,J.M. and Reed,S.G.
TITLE Compositions and methods for the treatment and diagnosis of breast
JOURNAL Patent: US 6225054-A 175 01-MAY-2001;
FEATURES
    source      1..484
                Location/Qualifiers
BASE COUNT 101 a 138 c 126 g 118 t 1 others
ORIGIN
Query Match      24.1%; Score 33.8; DB 6; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.46;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    25 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 78
Db 25 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 78

RESULT 7
LOCUS AX316838 484 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 175 from patent WO0190152.
ACCESSION AX316838
VERSION AX316838.1 GI:17899919
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1
TITLE Frudakis,T.N., Reed,S.G., Smith,J.M., Misher,L.E., Dillon,D.C.,
JOURNAL Rectler,M.W., Wang,A., Skeiky,Y.A., Harlocker,S.L. and Day,C.H.
TITLE Compositions and methods for the therapy and diagnosis of breast
JOURNAL Patent: WO 0190152-A 175 29-NOV-2001;
FEATURES
    source      1..484
                Location/Qualifiers
BASE COUNT 101 a 138 c 126 g 118 t 1 others
ORIGIN
Query Match      24.1%; Score 33.8; DB 6; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.46;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 11 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    1 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 64
    25 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 78
Db 25 GCATGCTCCGGCGCCATGCGCGGATTCATGAGCACTAACCCTTG 78

RESULT 8
LOCUS GGN132219 1271 bp mRNA linear PLN 23-NOV-1999
DEFINITION Gnetum gnetum mRNA for putative MADS domain transcription factor
ACCESSION GGN13.
    AJ132219

```


LSRGC2D1/C
LOCUS LSRGC2D1 5507 bp DNA linear PLN 12-JAN-1999
DEFINITION *Lactuca sativa* RGC2D pseudogene, partial sequence.
ACCESSION AF113953 AF072270
VERSION AF113953.1 GI:4150855
KEYWORDS
SEGMENT 1 of 2
SOURCE *Lactuca sativa*.
ORGANISM *Lactuca sativa*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asterales; Lactuceae; Lactuceae; Lactuca.
REFERENCE 1 (bases 1 to 5507)
AUTHORS Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z. and Michelmore, R.W.
TITLE The major resistance gene cluster in lettuce is highly duplicated and spans several megabases
JOURNAL Plant Cell 10 (11), 1817-1832 (1998)
MEDLINE 99030190
PUBMED 9811791
REFERENCE 2 (bases 1 to 5507)
AUTHORS Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S. and Michelmore, R.W.
TITLE Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection
JOURNAL Plant Cell 10 (11), 1833-1846 (1998)
MEDLINE 99030191
PUBMED 9811792
REFERENCE 3 (bases 1 to 5507)
AUTHORS Meyers, B.V., Shen, K.A., Rohani, P.J., Baul, B. and Michelmore, R.W.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) Vegetable Crops, UC Davis, 148 Asmundson, Davis, CA 95616, USA
FEATURES
source
1..5507
/organism="Lactuca sativa"
/db_xref="taxon:4236"
/map="Dm3 resistance gene cluster"
BASE COUNT 1798 a 940 c 1051 g 1717 t 1 others
ORIGIN
Query Match 23.9%; Score 33.4; DB 8; Length 5507;
Best Local Similarity 97.1%; Pred. No. 1.1;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCCGCGCCATGCGCGGATTGTCA 45
|||||
Db 5468 GCATGCTCCCGCGCCATGCGCGGATTGTAA 5434

RESULT 12
LOCUS HS296972 345 bp mRNA linear PRI 05-MAY-2001
DEFINITION *H. sapiens* mRNA for immunoglobulin heavy chain variable region (43-8; VH3, 3-15/DP-38).
ACCESSION Z96972.1 GI:2370221
VERSION Z96972.1
KEYWORDS antigen receptor; immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; variable region.
SOURCE human.
ORGANISM *Homo sapiens*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 345)
AUTHORS Domiatli-Saad, R. and Lipsky, P.E.
TITLE Staphylococcal enterotoxin A induces survival of VH3-expressing human B cells by binding to the VH region with low affinity
JOURNAL J. Immunol. 161 (3), 1257-1266 (1998)
MEDLINE 9866586
PUBMED 9686586
REFERENCE 2 (bases 1 to 345)
AUTHORS Domiatli-Saad, R.

TITLE Direct Submission
JOURNAL Submitted (23-JUN-1997) Domiatli-Saad R., Department of Internal Medicine, Harold C. Simmons Arthritis Research Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8884, USA
FEATURES
source
1..345
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/cell_type="B cell"
/tissue_type="peripheral blood"
/rearranged
1..345
/product="VH3, 3-15/DP-38"
V_region
BASE COUNT 75 a 83 c 116 g 70 t 1 others
ORIGIN
*Query Match 23.6%; Score 33; DB 9; Length 345;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCCGCGCCATGCGCGGATTGT 43
|||||
Db 330 GCATGCTCCCGCGCCATGCGCGGATTGT 298

RESULT 13
LOCUS AR210351 163 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 52 from patent US 6387657.
ACCESSION AR210351
VERSION AR210351.1 GI:21512560
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 163)
AUTHORS Botstein, D.A., Cohen, R.L., Goddard, A.D., Gurney, A.L., Hillan, K.J., Lawrence, D.A., Levine, A.J., Pennica, D., Roy, M.Ann. and Wood, W.I.
TITLE WISP polypeptides and nucleic acids encoding same
JOURNAL Patent: US 6387657-A 52 14-MAY-2002;
FEATURES
source
1..163
/organism="unknown"
BASE COUNT 36 a 46 c 45 g 36 t
ORIGIN
Query Match 23.3%; Score 32.6; DB 6; Length 163;
Best Local Similarity 89.7%; Pred. No. 1;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 GCATGCTCCCGCGCCATGCGCGGATTGTCA 49
|||||
Db 123 GCATGCTCCCGCGCCATGCGCGGATTATCACTAG 85

RESULT 14
LOCUS AC107595 155394 bp DNA linear HTG 13-JUL-2002
DEFINITION *Rattus norvegicus* clone CH230-273u5, *** SEQUENCING IN PROGRESS ***
ACCESSION AC107595
VERSION AC107595.3 GI:21737300
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM *Rattus norvegicus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 155394)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayala, M., Banks, T.,

Barbala, J., Benton, J., Bimage, K., Blankenburg, R., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carlton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Fratutz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokweto, S., Oguh, M., Okunolu, G.,
 Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Syatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, Y., Villalob, D., Vinson, R., Wang, C.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 2 (bases 1 to 155394)
 Worley, K.C.
 Direct Submission
 Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 155394)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:16701904.

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
 Center project name: GOWP
 Center clone name: CH230-27315

----- Summary Statistics -----
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 103821 bases at least Q40
 Consensus quality: 108962 bases at least Q30
 Consensus quality: 113213 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a working draft sequence. It currently
 * consists of 66 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1631: contig of 1631 bp in length
2	1632
3	1732: gap of unknown length
4	1732: contig of 1425 bp in length
5	3157: gap of unknown length
6	3257: contig of 1380 bp in length
7	4637: gap of unknown length
8	4737: contig of 1213 bp in length
9	5950: gap of unknown length
10	6050: contig of 1171 bp in length
11	7221: gap of unknown length
12	7321: contig of 1176 bp in length
13	8497: gap of unknown length
14	8597: gap of unknown length
15	8597: contig of 1134 bp in length
16	9731: gap of unknown length
17	9831: contig of 1689 bp in length
18	11520: gap of unknown length
19	11620: contig of 1658 bp in length
20	13278: gap of unknown length
21	13378: contig of 1607 bp in length
22	14985: gap of unknown length
23	15084: gap of unknown length
24	16292: contig of 1208 bp in length
25	16392: gap of unknown length
26	16393: contig of 1224 bp in length
27	17617: gap of unknown length
28	17617: contig of 1778 bp in length
29	19494: gap of unknown length
30	19595: contig of 2006 bp in length
31	21601: gap of unknown length
32	21701: contig of 1328 bp in length
33	23028: gap of unknown length
34	23029: contig of 1406 bp in length
35	23129: gap of unknown length
36	24535: contig of 2046 bp in length
37	24635: gap of unknown length
38	26680: gap of unknown length
39	26780: contig of 2266 bp in length
40	26781: gap of unknown length
41	29047: gap of unknown length
42	29147: contig of 1820 bp in length
43	30967: gap of unknown length
44	32514: contig of 1448 bp in length
45	32515: gap of unknown length
46	32615: contig of 2124 bp in length
47	34738: gap of unknown length
48	34739: contig of 1987 bp in length
49	34839: gap of unknown length
50	36825: contig of 1644 bp in length
51	36826: gap of unknown length
52	38569: gap of unknown length
53	38570: contig of 1445 bp in length
54	38670: gap of unknown length
55	40114: contig of 1974 bp in length
56	40115: gap of unknown length
57	40215: contig of 1261 bp in length
58	40215: gap of unknown length
59	42189: contig of 1039 bp in length
60	42289: gap of unknown length
61	43549: gap of unknown length
62	43550: contig of 1784 bp in length
63	43650: gap of unknown length
64	45433: gap of unknown length
65	45434: contig of 1653 bp in length
66	45534: gap of unknown length
67	47187: gap of unknown length
68	47286: contig of 1039 bp in length
69	48325: gap of unknown length
70	48425: contig of 1622 bp in length
71	50047: gap of unknown length
72	50147: contig of 2008 bp in length
73	50148: gap of unknown length
74	52155: contig of 2460 bp in length
75	52255: gap of unknown length
76	54715: contig of 1373 bp in length
77	54815: gap of unknown length
78	54816: contig of 1373 bp in length
79	56189: gap of unknown length
80	56289: contig of 2046 bp in length

REFERENCE 1 (bases 1 to 278)
AUTHORS Suoniemi, A., Schmidt, D. and Schulman, A. H.
TITLE BARE-1 insertion site preferences and evolutionary conservation of

Search completed: May 25, 2003, 19:41:04
Job time : 521.477 secs

2003-10-11 15:25

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:46:55 ; Search time 70.3404 Seconds

(without alignments)
4482.203 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
Sequence: 1 atagatatcagcatgcctccc.....gagctcgaatcgcctcata 140

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	100.0	140	21	AA245985
2	45	32.1	51	21	AA245973
3	39.6	28.3	477	21	AA245987
4	35.6	25.4	1949	22	AAK51591
5	35.4	25.3	1630	22	AAI57798
6	35.2	25.1	1222	20	AAV69558
7	34.8	24.9	2597	22	AAI58172
8	34.4	24.6	3186	22	AAI57799
9	34.2	24.4	163	20	AAK76514

10	34.2	24.4	163	20	AAK55630	5' Cloning site re
11	34	24.3	2064	24	ABN84401	Human Kv4.3 potass
12	34	24.3	2121	24	ABN84400	Human Kv4.3 potass
13	33.8	24.1	484	18	AAK83364	Breast cancer tumo
14	33.8	24.1	484	19	AAV68904	DNA molecule encod
15	33.8	24.1	484	21	AAK80887	Human breast tumo
16	33.8	24.1	484	21	ABK46777	Human breast tumo
17	33.8	24.1	484	24	AAK59733	Human breast tumo
18	33.4	23.9	5506	24	ABK67846	Breast tumour-spec
19	33.4	23.9	13339	19	AAV44234	Lettuce pest resis
20	33.2	23.7	711	22	ABA06302	Soy bean SCN/SCS r
21	33	23.6	3171	24	ABA93132	Infectious Bursal
22	32.6	23.3	163	20	AAK76515	WISP 3' cloning si
23	32.6	23.3	163	20	AAK55631	3' cloning site re
24	32.6	23.3	837	19	AAV49815	Human ADNF-III PCR
25	32.6	23.3	837	21	AAV49815	Sense PCR product
26	32.6	23.3	850	19	AAV49816	Human ADNF-III PCR
27	32.6	23.3	850	21	AAV49816	Antisense PCR prod
28	32.2	23.0	700	24	AAV0760	Pear pectin methyl
29	31.4	22.4	593	21	AAV7483	Physioacca america
30	31.4	22.4	713	22	ABA06304	Soy bean SCN/SCS r
31	31.4	22.4	831	24	ABN86270	S. vititatum silk p
32	31	22.1	215	22	AAH57722	Human breast tumo
33	31	22.1	420	19	AAV09802	DNA encoding the h
34	31	22.1	420	19	AAV09802	DNA encoding the l
35	31	22.1	482	23	AAV71699	DNA encoding novel
36	31	22.1	530	21	AAV47071	CDNA sequence of a
37	31	22.1	597	18	AAK83374	Breast cancer tumo
38	31	22.1	597	19	AAV68913	DNA molecule encod
39	31	22.1	597	21	AAK80897	Human breast tumo
40	31	22.1	597	24	ABK46787	Human breast tumo
41	31	22.1	597	24	AAK59743	Breast tumour-spec
42	31	22.1	919	21	AAK57434	Pythium venterpool
43	31	22.1	931	17	AAI14818	Human histiocyte-s
44	31	22.1	1205	18	AAI94303	ECM3 gene promoter
45	31	22.1	1265	22	AAI59000	Human polynucleoti

ALIGNMENTS

RESULT 1	AA245985	standard: cDNA; 140 Bp.
ID	AA245985	
XX	XX	
AC	AA245985	
XX	XX	
DT	25-APR-2000	(first entry)
XX	XX	
DE	Fragment of pJANG-deltaCys7Cys236 plasmid.	
XX	XX	
KW	Glutenin: seed storage protein; gluten; bread; pasta; noodle;	
KW	breakfast cereal; snack food; cake; pastry; flour based sauce; grain;	
KW	film; coating; adhesive; building material; packaging material; ss.	
OS	Synthetic.	
XX	XX	
OS	Hordeum vulgare.	
XX	XX	
FH	key	Location/Qualifiers
FT	CDS	45..107
FT		/tag= a
PN	WO200002914-A1.	
XX	XX	
PD	20-JAN-2000.	
XX	XX	
PF	12-JUL-1999:	99MO-AU00563.
XX	XX	
PR	10-JUL-1998:	98AU-0004604.
XX	XX	
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.	
PA	(GOOD-) GOODMAN FIELDER LTD.	
PA	(LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.	

```

XX
PI Appels R, Morell M, Bekes F, Tamas L;
XX
DR MPI: 2000-147597/13.
XX
DR P-PSDB: AAY54566.
XX
PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
PT films, coatings, packing materials, adhesives and building materials -
XX
PS Disclosure: Fig 1: 76pp: English.
XX
CC The present sequence represents a fragment of the pJANG-deltaCys70ys236
CC plasmid. This plasmid encodes a barley glutenin, and is used in the
CC course of the invention to produce modified glutenin. The specification
CC describes a method for producing a modified glutenin or seed storage
CC protein, by adding to the protein a domain that confers the ability to
CC incorporate into gluten, or to bind a ligand or other macromolecule. The
CC domain can be any domain that will bind ligands that may be useful in
CC food preparation or in food composition, e.g. a domain that binds lipids
CC or starch. The method is used for producing modified glutenins. Glutenins
CC and seed-storage protein are useful for preparing food products such as
CC leavened or unleavened breads, pasta, noodles, breakfast cereals, snack
CC foods, cakes, pastries and foods containing flour based sauces. Glutenins
CC and seed-storage protein are also useful for preparing nonfood products
CC such as films, coatings, adhesives, building materials and packaging
CC materials. Grain or parts of grain containing the modified glutenin and
CC seed-storage protein is useful for preparing food products. The modified
CC glutenins and seed-storage proteins are also useful as modifiers of
CC food products in food industry.
XX
SQ Sequence 140 BP: 34 A; 41 C; 36 G; 29 T; 0 other:
XX
Query Match 100.0%; Score 140; DB 21; Length 140;
Best Local Similarity 100.0%; Pred. No. 1,1e-38;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ATAGATATACAGCATGCTCCGCGCCCGCCATGCGGATTTGTCATGAGCACTAAACC 60
DB 1 ATAGATATACAGCATGCTCCGCGCCCGCCATGCGGATTTGTCATGAGCACTAAACC 60
XX
QY 61 CTTGCGAGCGTCCCCAACACAGCTTCATGATGAGTATGCTAGGATCCGGGTACC 120
DB 61 CTTGCGAGCGTCCCCAACACAGCTTCATGATGAGTATGCTAGGATCCGGGTACC 120
XX
QY 121 GAGCTCGAATTCGCCCTATA 140
DB 121 GAGCTCGAATTCGCCCTATA 140
XX
RESULT 2
AAZ45973/C
ID AAZ45973 standard; DNA: 51 BP.
XX
AC AAZ45973;
XX
DT 25-APR-2000 (first entry)
XX
DE PCR primer used to amplify a 447 bp fragment of C hordein gene.
XX
KW C hordein gene; glutenin; seed storage protein; gluten; bread; pasta;
KW noodle; breakfast cereal; snack food; cake; pastry; flour based sauce;
KW film; coating; adhesive; building material; packaging material; grain;
KW PCR primer; ss.
XX
OS Hordeum vulgare.
XX
PN WO200002914-A1.
XX
PD 20-JAN-2000.
XX
PF 12-JUL-1999; 99WO-AU00563.
XX
PR 10-JUL-1998; 98AU-0004604.

```

```

XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (GOOD-) GOODMAN FIELDER LTD.
PA (LIMA-) GRP LIMABRAIN PACIFIC PTY LTD.
XX
PI Appels R, Morell M, Bekes F, Tamas L;
XX
DR MPI: 2000-147597/13.
XX
PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
PT films, coatings, packing materials, adhesives and building materials -
XX
PS Disclosure: Page 11: 76pp: English.
XX
CC PCR primers AAZ45972-73 were used to amplify a 477 bp fragment
XX (approximately 2/3 of the whole gene) of the barley C hordein gene. This
XX gene encodes a glutenin, and was modified according to the method of
XX the invention. The specification describes a method for producing a
XX modified glutenin or seed storage protein, by adding to the protein a
XX domain that confers the ability to incorporate into gluten, or to bind
XX a ligand or other macromolecule. The domain can be any domain that will
XX bind ligands that may be useful in food preparation or in food
XX composition, e.g. a domain that binds lipids or starch. The method is
XX used for producing modified glutenins. Glutenins and seed-storage
XX protein are useful for preparing food products such as leavened or
XX unleavened breads, pasta, noodles, breakfast cereals, snack foods, cakes,
XX pastries and foods containing flour based sauces. Glutenins and
XX seed-storage protein are also useful for preparing nonfood products such
XX as films, coatings, adhesives, building materials and packaging
XX materials. Grain or parts of grain containing the modified glutenin and
XX seed-storage protein is useful for preparing food products. The modified
XX glutenins and seed-storage proteins are also useful as modifiers of food
XX products in food industry.
XX
SQ Sequence 51 BP: 11 A; 11 C; 15 G; 14 T; 0 other:
XX
Query Match 32.1%; Score 45; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 5.6e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 69 GTCCGCCAACAGCTTCATGATGATGAGTATGCTAGGATCC 113
DB 45 GTCCGCCAACAGCTTCATGATGATGAGTATGCTAGGATCC 1
XX
RESULT 3
AAZ45987
ID AAZ45987 standard; CDNA: 477 BP.
XX
AC AAZ45987;
XX
DT 25-APR-2000 (first entry)
XX
DE cDNA encoding a synthetically truncated C hordein protein (glutenin).
XX
KW C hordein gene; glutenin; seed storage protein; gluten; bread; pasta;
KW noodle; breakfast cereal; snack food; cake; pastry; flour based sauce;
KW film; coating; adhesive; building material; packaging material; grain;
KW ss.
XX
OS Hordeum vulgare.
XX
OS Synthetic.
XX
XX
FH Key 1..477 Location/Qualifiers
FT CDS
FT /tag= a
FT /transl_except= (pos: 58..60, aa: Gln)
FT /transl_except= (pos: 313..315, aa: Gln)
FT /transl_except= (pos: 316..318, aa: Gln)
FT /transl_except= (pos: 319..321, aa: Pro)
FT /transl_except= (pos: 322..324, aa: Phe)
FT /transl_except= (pos: 337..339, aa: Gln)
FT /transl_except= (pos: 340..342, aa: Gln)

```

FT FT /transl_except= (pos: 343..345, aa: Pro)
 FT /transl_except= (pos: 346..348, aa: Phe)
 FT /transl_except= (pos: 349..351, aa: Pro)
 FT /transl_except= (pos: 352..354, aa: Gln)
 FT /transl_except= (pos: 355..357, aa: Gln)
 FT /transl_except= (pos: 358..360, aa: Pro)
 XX
 XX WO200002914-A1.
 PN
 XX
 PD 20-JAN-2000.
 XX
 PE 12-JUL-1999: 99WO-AU00563.
 XX
 PR 10-JUL-1998: 98AU-0004604.
 XX
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (GOOD-) GOODMAN FIELDER LTD.
 PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
 XX
 PI Appels R, Morell M, Bekes F, Tamas L;
 DR WPI: 2000-147597/13.
 DR P-PSDB: AAY54568.
 XX
 PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
 PT films, coatings, packing materials, adhesives and building materials -
 XX
 PS Disclosure: Fig 3; 76pp; English.
 XX
 CC The present sequence encodes a synthetically truncated barley C
 CC hordein protein, which is a glutenin. The protein is designated
 CC CC-AMG-deltaCys75Cys236, and is modified, according to the method of
 CC the invention. The specification describes a method for producing a
 CC modified glutenin or seed storage protein, by adding to the protein a
 CC domain that confers the ability to incorporate into gluten, or to bind
 CC a ligand or other macromolecule. The domain can be any domain that will
 CC bind ligands that may be useful in food preparation or in food
 CC composition, e.g. a domain that binds lipids or starch. The method is
 CC used for producing modified glutenins. Glutenins and seed-storage protein
 CC are useful for preparing food products such as leavened or unleavened
 CC breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries
 CC and foods containing flour based sauces. Glutenins and seed-storage
 CC protein are also useful for preparing nonfood products such as films,
 CC coatings, adhesives, building materials and packaging materials. Grain
 CC or parts of grain containing the modified glutenin and seed-storage
 CC protein is useful for preparing food products. The modified glutenins and
 CC seed-storage proteins are also useful as modifiers of food products in
 CC food industry.
 CC
 XX
 SO Sequence 477 BP: 162 A; 186 C; 43 G; 86 T; 0 other:
 Query Match 28.3%; Score 39.6; DB 21: Length 477;
 Best Local Similarity 83.3%; Pred. No. 0.00076;
 Matches 45; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 OY 54 CTTAAACCTGCGACGCGCCGACCAAGCTCATGATGATGAGATGCTTAG 107
 DB 424 CAACACCTGACACGAGTCCGCCACACAGCTTCATGATGATGAGATGCTTAG 477

KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001: 2001WO-US04098.
 XX
 PR 03-FEB-2000: 2000US-0496914.
 PR 27-APR-2000: 2000US-0560875.
 PR 20-JUN-2000: 2000US-0598075.
 PR 19-JUL-2000: 2000US-0620325.
 PR 01-SEP-2000: 2000US-0654936.
 PR 15-SEP-2000: 2000US-0663561.
 PR 20-OCT-2000: 2000US-0693325.
 PR 30-NOV-2000: 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zheng J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PSDB: AAM78458.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 816-818; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52561), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SO Sequence 1949 BP: 484 A; 434 C; 517 G; 514 T; 0 other:
 Query Match 25.4%; Score 35.6; DB 22: Length 1949;
 Best Local Similarity 75.9%; Pred. No. 0.027;
 Matches 44; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 OY 11 GCATGCTCCCGCGCGCCGATGCGCGGATTCATGAGCAAGTAAACCTTGCAGC 68
 DB 10 GCATGCTCCCGCGCGCCGATGCGCGGATTTTCCGCGGATTCGATTCGTTGACC 67

RESULT 4
 AAK51591
 ID AAK51591 standard; cDNA: 1949 BP.
 XX
 AC AAK51591;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 136.
 XX
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;

RESULT 5
 AAI57798
 ID AAI57798 standard; cDNA: 1630 BP.
 XX
 AC AAI57798;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1.
 XX
 KW Human: neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW		chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW		leukemia; ss.
XX		
OS	Homo sapiens .	
XX		
PN	WO200153312-A1.	
PD	26-JUL-2001.	
XX		
PF	26-DEC-2000; 2000MO-US34263.	
XX		
PR	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0552317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
XX	29-NOV-2000; 2000US-0727344.	
PA	(HYSE-) HYSEQ INC.	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Dimanac RT;	
DR	WPI: 2001-442253/47.	
PT	P-PADB: AAM38642.	
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PS	such as central nervous system injuries -	
XX	Claim 1: SEQ ID NO 1; 10078bp; English.	
XX	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SQ	Sequence 1630 BP; 392 A; 422 C; 438 G; 378 T; 0 other:	
	Query Match 25.3%; Score 35.4; DB 22; Length 1630;	
	Best Local Similarity 79.2%; Pred. No. 0.03;	
	Matches 42; Conservative 0; Mismatches 11; Indels 0; Gaps 0	
OY	11 GCATGCTCGGGCGGCATGGCCGGGAGTTTCATGACGAACAACCCCT 63 	
Dn	10 GCATGCTCGGGCGGCATGGCCGGGAGTTTCGGGGTCAAGCATTTTCGTT 62	
RESULT 6		
ID	AAV69558/c	
XX	AAV69558 standard; DNA; 1222 BP.	
AC	AAV69558;	
XX		
DT	15-MAR-1999 (first entry)	
XX		
DE	Soil derived peptide synthase clone ps3 DNA.	
XX		
KW	Peptide synthase; soil; lichen; antibiotic biosynthesis; humus;	

KW	therapeutic; immunosuppressor; antitumour agent; pathogen;
KW	genetic diversity; ss.
XX	
OS	Unknown.
XX	
PN	W09853097-A2.
XX	
PD	26-NOV-1998.
XX	
PF	21-MAY-1998; 98W0-CA00488.
PR	22-MAY-1997; 97US-0861774.
XX	
PA	(TERR-) TERRAGEN DIVERSITY INC.
XX	
PI	Miao VPM, Seow KT, Waters B, Yap WH;
XX	
XX	WPI: 1999-070158/06.
DR	P-PSDB: AAM82710.
XX	
PT	New degenerate primers - used for recovering antibiotic biosynthetic
XX	DNA from soil/lichen material
PS	Claim 17; Page 80; 98pp; English.
XX	
CC	This sequence encodes a peptide synthase clone, isolated from soil.
CC	This protein is used in a method for the recovery of antibiotic
CC	biosynthetic DNA from humic materials or lichen. The PCR products of the
CC	invention have the potential to be used as therapeutic molecules
CC	including antibiotics, immunosuppressors and antitumour agents. The
CC	method allows access to the reservoir of genetic diversity in soil
CC	pathogenic micro-organisms. In order to find new antibiotics. It also
CC	allows access to novel biosynthetic genes/enzymes that can be used to
CC	produce antibiotics or produce specific compounds, enzymatically,
CC	in vitro.
XX	
SQ	Sequence 1222 BP; 272 A; 361 C; 319 G; 270 T; 0 other;
XX	
Query Match	25.1%; Score 35.2; DB 20; Length 1222;
Best Local Similarity	92.5%; Pred.No.0.033; 3; Indels 0; Gaps 0;
Matches 37; Conservative	0; Mismatches 3;
QY	11 GCATGCTCCCGCGCCATGCGCGCGGATTGTCATGAGG 50
DB	1176 GCATGCTCCCGCGCCATGCGCGCGGATTGCGACGACG 1137
XX	XX
RESULT 7	
AA158172/c	
ID	AA158172 standard; cDNA; 2597 BP.
XX	
AC	AA158172:
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 375.
XX	
XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukemia; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSEQ-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB: AAM39016.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1; SEQ ID NO 375; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 2597 BP; 733 A; 607 C; 690 G; 567 T; 0 other;
 XX
 XX
 XX Query Match 24.9%; Score 34.8; DB 22; Length 2597;
 XX Best Local Similarity 84.8%; Pred. No. 0.055; 7; Indels 0; Gaps 0;
 XX Matches 39; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCGCGGCGGATGTCATGAGGCACTA 56
 DB 2561 GCATGCTCCCGCGCCGCGGCGGATGTCATGAGGCACTA 2516
 RESULT 8
 ID AA157799 standard; CDNA; 3186 BP.
 AC AA157799;
 XX
 XX 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 2.
 XX
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 KW
 OS Homo sapiens.
 XX
 XX WO200153312-A1.
 XX
 XX PD 26-JUL-2001.
 XX

PF 26-DEC-2000; 2000MO-US34263.
 XX
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSEQ-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB: AAM38643.
 XX
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1; SEQ ID NO 2; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 3186 BP; 695 A; 1058 C; 894 G; 539 T; 0 other;
 XX
 XX
 XX Query Match 24.6%; Score 34.4; DB 22; Length 3186;
 XX Best Local Similarity 86.4%; Pred. No. 0.08;
 XX Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 11 GCATGCTCCCGCGCCGCGGCGGATGTCATGAGGCAAC 54
 DB 20 GCATGCTCCCGCGCCGCGGCGGATGTCATGAGGCAAC 63
 RESULT 9
 ID AAX76514 standard; DNA; 163 BP.
 AC AAX76514;
 XX
 XX 06-AUG-1999 (first entry)
 DE WISP 5' cloning site region SEQ ID NO:51.
 XX
 XX WNT-1 induced secreted protein; WISP-1; WISP-2; WISP-3; CTGF; tumour;
 KW connective tissue growth factor; cancer; melanoma; arteriosclerosis;
 KW leukaemia; lymphoid malignancy; haematopoiesis-related disorder;
 KW tissue-growth disorder; skin disorder; desmoplasia; fibrotic lesion;
 KW kidney disorder; bone-related disorder; osteoporosis; trauma; burn;
 KW connective tissue disorder; catabolic state; inflammation;
 KW testicular-related disorder; angiogenesis; immunological disorder; ss.
 KW
 OS Synthetic.
 XX

XX MO9921998-A1.
XX PD 06-MAY-1999.
XX PF 29-OCT-1998; 98WO-US22991.
XX 14-APR-1998; 98US-0081695.
PR 29-OCT-1997; 97US-0063704.
PR 03-FEB-1998; 98US-0073612.
XX (GETH) GENENTECH INC.
XX Botstein DA, Cohen RL, Goddard A, Gurney AL, Hillan K;
PI Lawrence DA, Levine AJ, Pennica D, Roy MA, Wood WI;
XX WPI: 1999-337420/28.
DR New isolated Wnt-1 induced secreted polypeptides, WISP-1, 2 and 3
PT
XX
XX Example 1; Page 203; 284pp; English.
XX
XX The present invention describes Wnt-1 induced secreted polypeptides,
CC WISP-1, 2 and 3. The novel WISP polypeptides, designated WISP-1, WISP-2
CC and WISP-3 have homology to connective tissue growth factor (CTGF).
CC Products from the present invention can be used to treat WISP-related
CC disorders such as breast, ovarian, and colon cancer or melanoma. The
CC products can be used to treat arteriosclerosis. The products can also be
CC used to treat other diseases e.g. benign and malignant tumours,
CC leukaemia and lymphoid malignancies, neuronal, glial, astrocytal,
CC hypothalamic and other glandular, macrophagal, epithelial, stromal, and
CC blastocoelec disorders, haematopoiesis-related disorders, tissue-growth
CC disorders, skin disorders, desmoplasia, fibrotic lesions, kidney
CC disorders, bone-related disorders such as osteoporosis, trauma such as
CC burns, incisions, and other wounds, connective tissue disorders,
CC catabolic states, testicular-related disorders, and inflammatory,
CC angiogenic and immunologic disorders including arteriosclerosis. The
CC products can also be used for detection and diagnosis especially of
CC individuals with neoplastic cell growth or proliferation. The products
CC can be used in the production of transgenic or knock-out animals.
CC Antibodies can be used to induce death in WISP-1, 2 or 3 overexpressing
CC cells.
XX
XX Sequence 163 BP; 35 A; 45 C; 46 G; 37 T; 0 other;
SQ
XX
XX Query Match 24.4%; Score 34.2; DB 20; Length 163;
XX Best Local Similarity 92.3%; Pred. No. 0.041;
XX Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 11 GCATGCTCCGGCGCCGCAATGGCGGGGATTTCTCATGAG 49
XX 11
Db 41 GCATGCTCCGGCGCCGCAATGGCGGGGATTTCTCATGAG 79
XX 41
RESULT 10
ID AAX55630
AAX55630 standard; DNA; 163 BP.
XX
XX AAX55630:
XX
XX 26-JUL-1999 (first entry)
XX
XX 5' cloning site region of plasmid pGEM-T.
DE
XX
XX Human: murine; clone 65; clone 320; Wnt; arteriosclerosis; malignancy;
KW melanoma; cancer; breast; ovary; colon; tumor; cardiac; renal;
KW inflammation; angiogenesis; immunological disorder; ss.
XX
XX Synthetic.
XX
XX WO9921999-A2.
XX
XX PD 06-MAY-1999.

```

PF 29-OCT-1998: 98MO-US22992.
XX 04-FEB-1998: 98US-0073612.
PR 29-OCT-1997: 97US-0063704.
XX
PA (GETH ) GENENTECH INC.
XX
PI Levine AJ, Pennica D;
XX
DR WPI; 1999-326705/27.
XX
PT Human and murine genes induced by Wnt and derived polypeptides
PS
XX Examples: Fig 9; 114pp; English.
XX
CC The invention relates to human and murine cDNA clones 65 and 320,
CC encoding polypeptides that are induced by Wnt. Clone 65 and 320
CC polypeptides, and their antagonists, are used to treat disorders
CC associated with these polypeptides, e.g. arteriosclerosis, malignancies
CC (particularly melanoma or cancer of the breast, ovary and colon), but
CC also benign tumors, cardiac, renal, catabolic or many other types of
CC disease, inflammation, angiogenesis and immunological disorders, more
CC generally to induce death of Wnt-induced cells. Antibodies against the
CC polypeptides are used as therapeutic antagonists and as diagnostic
CC immunoassay or affinity-purification reagents. Clone 65 and 320 nucleic
CC acids are used as hybridization probes or primers (to detect related
CC sequences); for chromosome and gene mapping; to generate antisense
CC sequences; for recombinant polypeptide production and to generate
CC transgenic or 'knockout' animals (for development of, and screening for,
CC drugs). The polypeptides are used to raise (or purify) specific
CC antibodies; as immunoassay reagents and in drug screens.
XX
SQ Sequence 163 BP; 35 A; 45 C; 46 G; 37 T; 0 other;
XX
Query Match 24.4%; Score 34.2; DB 20; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.041;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
OY 11 GCATGCTCCGGCGGCATGGCGCGGATTTGTCATGAG 49
| | | | | | | | | | | | | | | | | | | | | |
DB 41 GCATGCTCCGGCGGCATGGCGCGGATTTATCACTAG 79

RESULT 11
ABN84401/C
ID ABN84401 standard; cDNA; 2064 BP.
XX
XX ABN84401;
XX
XX AC
XX DT 01-OCT-2002 (first entry)
XX
XX DE Human Kv4.3 potassium channel (short form) cDNA.
XX
XX KW Kv4.3; potassium channel; human; Alzheimer's disease;
XX heart disease; nontoxic; neuroprotective; cardiac; gene therapy;
XX gene; ss.
XX
XX OS Homo sapiens.
XX
XX FH
XX FT Key Location/Qualifiers
XX FT CDS 73..1983
XX FT /*tag= a
XX FT /*product= "Kv4.3"
XX
XX PN US6395477-B1.
XX
XX PD 28-MAY-2002.
XX
XX PF 23-OCT-1998; 98US-0178109.
XX
XX PR 23-OCT-1998; 98US-0178109.
XX
XX
XX (AMHP ) AMERICAN HOME PROD CORP.

```

XX PI Cockett MI, Dilks DW, Ling HC, Sokol PT;
XX DR WPI: 2002-556093/59.
XX DR P-PSDB: ABB79585.
XX PT New isolated polynucleotide encoding human Kv4.3 potassium channel
XX PT polypeptide, useful as probe in a diagnostic method for detecting
XX PT nucleic acid encoding human Kv4.3, and for treating Alzheimer's and
XX PT heart diseases -
XX PS Claim 7; Column 21-22; 19pp; English.
XX CC The present sequence is that of cDNA encoding the short isoform
XX CC of novel human potassium channel Kv4.3. 2 isoforms of human Kv4.3
XX CC have been identified. One form is full-length (hKv4.3 long) while
XX CC the second form has a deletion of 19 amino acids in the carboxy
XX CC domain after the predicted sixth transmembrane domain (hKv4.3 short).
XX CC Human heart primarily expresses hKv4.3 long, whereas human brain
XX CC contains both forms. To obtain the present hKv4.3 short cDNA, PCR
XX CC amplification was performed using primers that flanked the 57 bp
XX CC insert in hKv4.3 long. The invention provides Kv4.3 polypeptides,
XX CC polynucleotides, and methods for producing these polynucleotides.
XX CC The Kv4.3 polypeptides and polynucleotides are useful in the
XX CC diagnosis, treatment and screening of human diseases relating to an
XX CC excess or deficiency of hKv4.3 activity, including Alzheimer's
XX CC disease and heart disease.
XX SO Sequence 2064 BP; 446 A; 678 C; 545 G; 395 T; 0 other;
Query Match 24.3%; Score 34; DB 24; Length 2064;
Best Local Similarity 88.1%; Pred. No. 0.098;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
DB 2015 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 1974
RESULT 12
ABN84400/c
ID ABN84400 standard; cDNA: 2121 BP.
XX AC ABN84400;
XX DT 01-OCT-2002 (first entry)
XX DE Human Kv4.3 potassium channel (long form) cDNA.
XX KW Kv4.3; potassium channel; human; Alzheimer's disease;
XX KW heart disease; nootropic; neuroprotective; cardiac; gene therapy;
XX KW gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 73..2040
XX FT /*tag= a
XX FT //product= "Kv4.3"
XX FT US6395477-B1.
XX PN 28-MAY-2002.
XX PD 23-OCT-1998; 98US-0178109.
XX PR 23-OCT-1998; 98US-0178109.
XX PA (AMHP) AMERICAN HOME PROD CORP.
XX PI Cockett MI, Dilks DW, Ling HC, Sokol PT;
XX DR WPI: 2002-556093/59.

XX DR P-PSDB: ABB79584.
XX PT New isolated polynucleotide encoding human Kv4.3 potassium channel
XX PT polypeptide, useful as probe in a diagnostic method for detecting
XX PT nucleic acid encoding human Kv4.3, and for treating Alzheimer's and
XX PT heart diseases -
XX PS Claim 5; Column 15-18; 19pp; English.
XX CC The present sequence is that of cDNA encoding the long isoform of
XX CC novel human potassium channel Kv4.3. To obtain the cDNA,
XX CC oligonucleotides based on the published rat sequence were used to
XX CC screen a whole heart cDNA library. A 511 bp fragment was obtained
XX CC and used as a probe to rescreen the library. The resulting clones
XX CC lacked the extreme 5' and 3' coding regions, and 5' and 3' RACE
XX CC was therefore used to amplify these sequences from a human
XX CC brainstem cDNA library. 2 isoforms of human Kv4.3 were identified.
XX CC One form is full-length (hKv4.3 long) while the second form has a
XX CC deletion of 19 amino acids in the carboxy domain after the
XX CC predicted sixth transmembrane domain (hKv4.3 short). Human heart
XX CC primarily expresses hKv4.3 long, whereas human brain contains both
XX CC forms. The invention provides Kv4.3 polypeptides, polynucleotides,
XX CC and methods for producing these polynucleotides. The Kv4.3
XX CC polypeptides and polynucleotides are useful in the diagnosis,
XX CC treatment and screening of human diseases relating to an excess or
XX CC deficiency of hKv4.3 activity, including Alzheimer's disease and
XX CC heart disease.
XX SO Sequence 2121 BP; 457 A; 694 C; 557 G; 413 T; 0 other;
Query Match 24.3%; Score 34; DB 24; Length 2121;
Best Local Similarity 88.1%; Pred. No. 0.099;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
DB 2072 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 2031
RESULT 13
AAX83364
ID AAX83364 standard; cDNA: 484 BP.
XX AC AAX83364;
XX DT 31-AUG-1999 (first entry)
XX DE Breast cancer tumour specific clone #118.
XX KW Breast cancer; tumour; gene expression; genome; diagnosis; mammal;
XX KW human endogenous retrovirus; vaccine; ss.
XX OS Homo sapiens.
XX PN W09725426-A2.
XX PD 17-JUL-1997.
XX PF 10-JAN-1997; 97WO-US00485.
XX PR 20-AUG-1996; 96US-0700014.
XX PR 11-JAN-1996; 96US-0585392.
XX PA (CORI-) CORIXA CORP.
XX PI Frudakis TN, Reed SG, Smith JM.
XX DR WPI: 1997-372865/34.
XX PT Breast cancer-related DNA from retrovirus antigen (s) - useful for
XX PT diagnosis and treatment of breast cancer
XX PS Claim 1; Page 150-151; 221pp; English.

QY 11 GCATGCTCCCGGGCCCATGGCCGGGATTGTTCATGAGGCACTAAACCCCTG 64

Search completed: May 25, 2003, 19:02:42
Job time : 71.3404 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:55:25 ; Search time 15.2026 Seconds

(without alignments)
2824.173 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
1 atagaatacagatgcctcc.....gagctcgaatgcctcata 140

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35.2	25.1	1222	4 US-08-861-774E-83	Sequence 83, Appl
C 2	34.2	24.4	163	4 US-09-182-145-51	Sequence 51, Appl
C 3	34	24.3	2054	4 US-09-178-109-3	Sequence 3, Appl
C 4	34	24.3	2121	4 US-09-178-109-1	Sequence 1, Appl
C 5	33.8	24.1	484	4 US-08-991-789A-175	Sequence 175, App
C 6	33.8	24.1	484	4 US-09-062-451-175	Sequence 175, App
C 7	33.8	24.1	484	4 US-09-598-326-175	Sequence 175, App
C 8	33.4	23.9	5506	4 US-09-004-838-93	Sequence 93, Appl
C 9	32.6	23.3	163	4 US-09-182-145-52	Sequence 52, Appl
C 10	31	22.1	419	4 US-09-214-095D-97	Sequence 97, Appl
C 11	31	22.1	420	4 US-09-214-095D-111	Sequence 111, App
C 12	31	22.1	597	4 US-08-991-789A-185	Sequence 185, App
C 13	31	22.1	597	4 US-09-062-451-185	Sequence 185, App
C 14	31	22.1	597	4 US-09-598-326-185	Sequence 185, App
C 15	30.8	22.0	500	4 US-08-991-789A-167	Sequence 167, App
C 16	30.8	22.0	500	4 US-09-062-451-167	Sequence 167, App
C 17	30.8	22.0	500	4 US-09-598-326-167	Sequence 167, App
C 18	30.4	21.7	488	4 US-08-991-789A-180	Sequence 180, App
C 19	30.4	21.7	488	4 US-09-062-451-180	Sequence 180, App
C 20	30.4	21.7	488	4 US-09-598-326-180	Sequence 180, App
C 21	30.4	21.7	507	4 US-08-991-789A-182	Sequence 182, App
C 22	30.4	21.7	507	4 US-09-062-451-182	Sequence 182, App
C 23	30.4	21.7	507	4 US-09-598-326-182	Sequence 182, App
C 24	30	21.4	360	4 US-09-214-095D-91	Sequence 91, Appl
C 25	30	21.4	5506	4 US-09-004-838-93	Sequence 93, Appl
C 26	29.8	21.3	5829	4 US-09-605-785-473	Sequence 473, App
C 27	29.8	21.3	5829	4 US-09-439-513-473	Sequence 473, App

28	29.4	21.0	597	4 US-08-991-789A-186	Sequence 186, App
29	29.4	21.0	597	4 US-09-062-451-186	Sequence 186, App
30	29.4	21.0	597	4 US-09-598-326-186	Sequence 186, App
31	29	20.7	93	4 US-08-986-659B-15	Sequence 15, Appl
32	29	20.7	165	2 US-08-484-956-49	Sequence 49, Appl
33	29	20.7	165	2 US-08-757-653-49	Sequence 49, Appl
34	29	20.7	165	2 US-08-520-946-49	Sequence 49, Appl
C 35	29	20.7	206	1 US-08-254-359A-32	Sequence 32, Appl
C 36	29	20.7	206	1 US-08-471-066B-32	Sequence 32, Appl
C 37	29	20.7	206	2 US-08-484-956-32	Sequence 32, Appl
C 38	29	20.7	206	2 US-08-484-956-50	Sequence 50, Appl
C 39	29	20.7	206	2 US-08-757-653-32	Sequence 32, Appl
C 40	29	20.7	206	2 US-08-757-653-50	Sequence 30, Appl
C 41	29	20.7	206	2 US-08-599-491-32	Sequence 32, Appl
C 42	29	20.7	206	2 US-08-756-386-32	Sequence 32, Appl
C 43	29	20.7	206	2 US-08-823-516-27	Sequence 27, Appl
C 44	29	20.7	206	3 US-08-682-853A-32	Sequence 32, Appl
C 45	29	20.7	206	3 US-08-759-038-32	Sequence 32, Appl

ALIGNMENTS

```

RESULT 1
US-08-861-774E-83/c
; Sequence 83, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 1222
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps7
US-08-861-774E-83

Query Match      25.1%; Score 35.2; DB 4; Length 1222;
Best Local Similarity 92.5%; Pred. No. 0.0029;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      11 GCATGCTCCCGCCGCCGATGCGCGGATGTCATGAGG 50
DB      1176 GCATGCTCCCGCCGCCGATGCGCGGATGTCGACGAGG 1137

RESULT 2
US-09-182-145-51
; Sequence 51, Application US/09182145B
; Patent No. 6387657
; GENERAL INFORMATION:
; APPLICANT: Bolstein, David A.
; APPLICANT: Cohen, Robert
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Lawrence, David J.
; APPLICANT: Levine, Arnold J.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: MSP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: P1176R2

```

```
; CURRENT APPLICATION NUMBER: US/09/182,145B
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: US 60/063,704
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: US 60/073,612
; EARLIER FILING DATE: 1998-02-04
; EARLIER APPLICATION NUMBER: US 60/081,695
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 51
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1-163
; OTHER INFORMATION: Sequence is synthesized
; Patent No. 6387657
US-09-182-145-51
```

```
Query Match          24.4%; Score 34.2; DB 4; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.0038;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 49
         |||||||
Db      41 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 79
```

```
RESULT 3
US-09-178-109-3/c
; Sequence 3, Application US/09178109
; Patent No. 6395477
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dilks, Daniel W.
; APPLICANT: Chang Ling, Hwai-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; TITLE OF INVENTION: Polypeptides and Uses Therefor
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/09/178,109
; CURRENT FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: human
US-09-178-109-3
```

```
Query Match          24.3%; Score 34; DB 4; Length 2064;
Best Local Similarity 88.1%; Pred. No. 0.009;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
         |||||||
Db      2015 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 1974
```

```
RESULT 4
US-09-178-109-1/c
; Sequence 1, Application US/09178109
; Patent No. 6395477
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dilks, Daniel W.
; APPLICANT: Chang Ling, Hwai-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; TITLE OF INVENTION: Polypeptides and Uses Therefor
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/09/178,109
```

```
; CURRENT FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2121
; TYPE: DNA
; ORGANISM: human
US-09-178-109-1
```

```
Query Match          24.3%; Score 34; DB 4; Length 2121;
Best Local Similarity 88.1%; Pred. No. 0.0091;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 52
         |||||||
Db      2072 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCA 2031
```

```
RESULT 5
US-08-991-789A-175
; Sequence 175, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
```

```
APPLICANT: Frudakis, Tony N.
```

```
Smith, John M.
```

```
Reed, Steven G.
```

```
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
```

```
NUMBER OF SEQUENCE: 292
```

```
CORRESPONDENCE ADDRESS:
```

```
ADDRESS: Seed IP Law Group
```

```
STREET: 701 Fifth Avenue, Suite 6300
```

```
CITY: Seattle
```

```
STATE: Washington
```

```
COUNTRY: USA
```

```
ZIP: 98104-7092
```

```
COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/991,789A
```

```
FILING DATE: 11-Dec-1997
```

```
CLASSIFICATION: <unknown>
```

```
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Potter, Jane E. R.
```

```
REGISTRATION NUMBER: 33,332
```

```
REFERENCE/DOCKET NUMBER: 210121.419C3
```

```
TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: (206) 622-4900
```

```
TELEFAX: (206) 682-6031
```

```
INFORMATION FOR SEQ ID NO: 175:
```

```
SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 484 base pairs
```

```
TYPE: nucleic acid
```

```
STRANDEDNESS: single
```

```
TOPOLOGY: linear
```

```
SEQUENCE DESCRIPTION: SEQ ID NO: 175:
```

```
US-08-991-789A-175
```

```
Query Match          24.1%; Score 33.8; DB 4; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.0071;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCAACTAAACCTTG 64
         |||||||
Db      25 GCATGCTCCCGCGCCGATGCGCGGATTCATGAGCAACTAAACCTTG 78
```

```
RESULT 6
US-09-062-451-175
; Sequence 175, Application US/09062451
```


LOCATION: 1..5506
OTHER INFORMATION: /note= "R32D"
US-09-004-838-93

Query Match 23.9%; Score 33.4; DB 4; Length 5506;
Best Local Similarity 97.1%; Pred. No. 0.019;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCCATGCGCGGATTTGCA 45
|||||
DB 5468 GCATGCTCCGCGCCCATGCGCGGATTTGTA 5434

RESULT 9
US-09-182-145-52/C

Sequence 52, Application US/09182145B

Patent No. 6387657

GENERAL INFORMATION:

APPLICANT: Botstein, David A.

APPLICANT: Cohen, Robert

APPLICANT: Goddard, Audrey

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Lawrence, David A.

APPLICANT: Levine, Arnold J.

APPLICANT: Pennica, Diane

APPLICANT: Roy, Margaret Ann

APPLICANT: Wood, William I.

TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

FILE REFERENCE: P1176R2

CURRENT APPLICATION NUMBER: US/09/182,145B

EARLIER FILING DATE: 1998-10-29

EARLIER FILING DATE: 1997-10-29

EARLIER APPLICATION NUMBER: US 60/073,612

EARLIER FILING DATE: 1998-02-04

EARLIER APPLICATION NUMBER: US 60/081,695

NUMBER OF SEQ ID NOS: 156

SEQ ID NO 52

LENGTH: 163

TYPE: DNA

ORGANISM: Artificial sequence

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1-163

OTHER INFORMATION: Sequence is synthesized

Patent No. 6387657

US-09-182-145-52

Query Match 23.3%; Score 32.6; DB 4; Length 163;

Best Local Similarity 89.7%; Pred. No. 0.014;

Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCCATGCGCGGATTTGCA 49
|||||

DB 123 GCATGCTCCGCGCCCATGCGCGGATTTGTA 85

RESULT 10
US-09-214-095D-97

Sequence 97, Application US/09214095D

Patent No. 6280987

GENERAL INFORMATION:

APPLICANT: Landry, Donald

TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY

FILE REFERENCE: 51400-A-PCT-US

CURRENT APPLICATION NUMBER: US/09/214,095D

CURRENT FILING DATE: 1999-07-19

NUMBER OF SEQ ID NOS: 121

SOFTWARE: PatentIn version 3.0

SEQ ID NO 97

LENGTH: 419

TYPE: DNA
ORGANISM: Murine
US-09-214-095D-97

Query Match 22.1%; Score 31; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCCATGCGCGGATTT 41
|||||

DB 18 GCATGCTCCGCGCCCATGCGCGGATTT 48

RESULT 11
US-09-214-095D-111

Sequence 111, Application US/09214095D

Patent No. 6280987

GENERAL INFORMATION:

APPLICANT: Landry, Donald

TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY

FILE REFERENCE: 51400-A-PCT-US

CURRENT APPLICATION NUMBER: US/09/214,095D

CURRENT FILING DATE: 1999-07-19

NUMBER OF SEQ ID NOS: 121

SOFTWARE: PatentIn version 3.0

SEQ ID NO 111

LENGTH: 420

TYPE: DNA

ORGANISM: Murine

FEATURE:

NAME/KEY: V-segment

LOCATION: (1)..(403)

OTHER INFORMATION: n at any position represents any nucleotide including c,g,t,a,

US-09-214-095D-111

Query Match 22.1%; Score 31; DB 4; Length 420;

Best Local Similarity 100.0%; Pred. No. 0.067;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGCGCCCATGCGCGGATTT 41
|||||

DB 17 GCATGCTCCGCGCCCATGCGCGGATTT 47

RESULT 12
US-08-991-789A-185

Sequence 185, Application US/08991789A

Patent No. 6225054

GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.

Smith, John M.

Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESS: Seed IP Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A

FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-991-789A-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCCGCATGCGCGGATT 42

RESULT 13
US-09-062-451-185
Sequence 185 Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
REFERENCE/DOCKET NUMBER: 33.332
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCCGCATGCGCGGATT 42

RESULT 14
US-09-598-326-185
Sequence 185 Application US/09598326
Patent No. 6423496
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
REFERENCE/DOCKET NUMBER: 33.332
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 247
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group PLLC
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33.332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4800
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-09-598-326-185

Query Match 22.1%; Score 31; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCCGCATGCGCGGATT 41
11 |||||
DB 12 GCATGCTCCGGCCCGCATGCGCGGATT 42

RESULT 15
US-08-991-789A-167
Sequence 167 Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
ATTORNEY/AGENT INFORMATION:
NAME: Smith, John M.
REFERENCE/DOCKET NUMBER: 31,392
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
MEDIUM TYPE: floppy disk

Wed May 28 09:17:34 2003

us-09-743-533-14.rni

Page 6

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Pottery, Jane E., R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.41903
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 167:
US-08-991-789A-167

```

	Query Match	92.0%	Score 30.8	DB 4	Length 500
	Best Local Similarity	94.1%	Pred. No.	0.082	
Matches	32; Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
OY	11 GCATGCTCCCGGCCCATGGCCCCGSGATTTC	44			
D6	7 GCATGCTCCCGGCCCATGGCCCCGSGATTGAC	40			

Search completed: May 25, 2003, 20:18:29
Job time : 17.2026 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:57:50 : Search time 44.2464 Seconds
(without alignments)
4178.078 Million cell updates/sec

Title: US-09-743-533-14

Perfect score: 140
Sequence: 1 atagaatacagatgctcccc.....gagctcgaattgcctcata 140

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/pdata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/pdata/2/pubpna/PC7_NEW_PUB.seq.*
3: /cgn2_6/pdata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/pdata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/pdata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/pdata/2/pubpna/PC7US_PUBCOMB.seq.*
7: /cgn2_6/pdata/2/pubpna/PC7US_PUBCOMB.seq.*
8: /cgn2_6/pdata/2/pubpna/US08_NEW_PUB.seq.*
9: /cgn2_6/pdata/2/pubpna/US08_PUBCOMB.seq.*
10: /cgn2_6/pdata/2/pubpna/US09_NEW_PUB.seq.*
11: /cgn2_6/pdata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/pdata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/pdata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/pdata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.4	25.3	1630	9 US-10-098-841-1	Sequence 1, Appl
2	35.2	25.1	1222	10 US-09-924-256A-83	Sequence 83, Appl
3	34.4	24.6	3186	9 US-10-098-841-2	Sequence 2, Appl
4	34.2	24.4	163	9 US-10-112-267-51	Sequence 51, Appl
5	34	24.3	2064	12 US-10-062-879-3	Sequence 3, Appl
6	34	24.3	2121	12 US-10-062-879-1	Sequence 1, Appl
7	33.8	24.1	484	9 US-09-924-400-175	Sequence 175, App
8	33.8	24.1	484	10 US-09-810-936-175	Sequence 175, App
9	33.8	24.1	484	10 US-09-429-755-175	Sequence 175, App
10	33.2	23.7	711	10 US-09-772-134B-80	Sequence 80, Appl
11	32.6	23.3	163	9 US-10-112-267-52	Sequence 52, Appl
12	31.4	22.4	669	9 US-10-043-875-841	Sequence 841, App
13	31.4	22.4	713	10 US-09-772-134B-82	Sequence 82, Appl
14	31.4	22.4	831	9 US-09-694-916-6	Sequence 6, Appl
15	31	22.1	215	10 US-09-778-320-250	Sequence 250, App
16	31	22.1	215	10 US-09-910-689-250	Sequence 250, App
17	31	22.1	215	12 US-10-010-742-250	Sequence 97, Appl
18	31	22.1	419	9 US-09-940-727B-97	Sequence 111, App
19	31	22.1	420	9 US-09-940-727B-111	

20	31	22.1	581	10 US-09-728-952-25	Sequence 25, Appl
21	31	22.1	597	9 US-09-924-400-185	Sequence 185, App
22	31	22.1	597	10 US-09-810-936-185	Sequence 185, App
23	31	22.1	597	10 US-09-429-755-185	Sequence 73, Appl
24	31	22.1	1680	9 US-09-298-523B-73	Sequence 38, Appl
25	31	22.1	1766	9 US-09-298-523B-74	Sequence 167, App
26	31	22.1	2714	10 US-09-835-996A-38	Sequence 167, App
27	30.8	22.0	500	9 US-09-924-400-167	Sequence 167, App
28	30.8	22.0	500	10 US-09-810-936-167	Sequence 167, App
29	30.8	22.0	500	10 US-09-429-755-167	Sequence 180, App
30	30.4	21.7	488	10 US-09-924-400-180	Sequence 180, App
31	30.4	21.7	488	10 US-09-810-936-180	Sequence 180, App
32	30.4	21.7	488	10 US-09-429-755-180	Sequence 182, App
33	30.4	21.7	507	9 US-09-924-400-182	Sequence 182, App
34	30.4	21.7	507	10 US-09-810-936-182	Sequence 182, App
35	30.4	21.7	507	10 US-09-429-755-182	Sequence 91, Appl
36	30	21.4	360	9 US-09-940-727B-91	Sequence 24, Appl
37	30	21.4	17458	9 US-10-055-001A-25	Sequence 26, Appl
38	30	21.4	17476	9 US-10-055-001A-26	Sequence 26, Appl
39	30	21.4	17681	9 US-10-055-001A-26	Sequence 26, Appl
40	30	21.4	17682	9 US-10-055-001A-23	Sequence 13, Appl
41	30	21.4	18691	9 US-10-055-001A-13	Sequence 473, App
42	29.8	21.3	5829	9 US-10-012-896-473	Sequence 473, App
43	29.8	21.3	5829	9 US-09-895-793-473	Sequence 472, App
44	29.8	21.3	5829	9 US-09-895-814-473	Sequence 473, App
45	29.8	21.3	5829	9 US-10-010-940-473	

ALIGNMENTS

RESULT 1
US-10-098-841-1
Sequence 1, Application US/10098841
Publication No. US20020197679A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungding
APPLICANT: Wang, JianRui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyang
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Duntui
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/10/098, 841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598, 042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552, 317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488, 725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 1
LENGTH: 1630
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (261)..(1052)
US-10-098-841-1

Query Match 25.3%; Score 35.4; DB 9; Length 1630;
Best Local Similarity 79.2%; Pred. No. 0.0038;
Matches 42; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 11 GCATGCTCCGGCCGCGCATGGCGGGGATGTCATGACCACTAACCCCTT 63
|||||
10 GCATGCTCCGGCCGCGCATGGCGGGGATTTTCCGGGTGACGATTCGTT 62

RESULT 2
US-09-924-256A-83/c
; Sequence 83, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 1222
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone ps7
US-09-924-256A-83

Query Match 25.1%; Score 35.2; DB 10; Length 1222;
Best Local Similarity 92.5%; Pred. No. 0.0043;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 11 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACG 50
|||||
Db 1176 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACG 1137

RESULT 3
US-10-098-841-2
; Sequence 2, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehtman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20

; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pl_FL_genes Version 1.0
; SEQ ID NO 2
; LENGTH: 3186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1848)
US-10-098-841-2

Query Match 24.6%; Score 34.4; DB 9; Length 3186;
Best Local Similarity 86.4%; Pred. No. 0.01;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 11 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACGAC 54
|||||
Db 20 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACGAC 63

RESULT 4
US-10-112-267-51
; Sequence 51, Application US/10112267
; Publication No. US2003006678A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David A.
; APPLICANT: Cohen, Robert
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Lawrence, David A.
; APPLICANT: Levine, Arnold J.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: P1176R2
; CURRENT APPLICATION NUMBER: US/10/112,267
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/182,145B
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/063,704
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/073,612
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/081,695
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 51
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1-163
; OTHER INFORMATION: Sequence is synthesized
US-10-112-267-51

Query Match 24.4%; Score 34.2; DB 9; Length 163;
Best Local Similarity 92.3%; Pred. No. 0.0063;
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 11 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACG 49
|||||
Db 41 GCATGCTCCGGCCGCGCATGGCGGGGATTCATGACG 79

RESULT 5
US-10-062-879-3/c

```
; Sequence 3, Application US/10062879
; Patent No. US20020127649A1
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dिल्s, Daniel W.
; APPLICANT: Chang Ling, Hual-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/10/062,879
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/178,109
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: human
US-10-062-879-3
```

```
Query Match          24.3%; Score 34; DB 12; Length 2064;
Best Local Similarity 88.1%; Pred. No. 0.013;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATGTCATGAGCA 52
Db      2015 GCATGCTCCCGCGCCGATGCGCGGATGTTTACAAGACA 1974
```

RESULT 6

```
US-10-062-879-1/c
; Sequence 1, Application US/10062879
; Patent No. US20020127649A1
; GENERAL INFORMATION:
; APPLICANT: Cockett, Mark I.
; APPLICANT: Dिल्s, Daniel W.
; APPLICANT: Chang Ling, Hual-Ping
; APPLICANT: Sokol, Patricia T.
; TITLE OF INVENTION: Human Potassium Channel Polynucleotides and
; FILE REFERENCE: ahp-98089
; CURRENT APPLICATION NUMBER: US/10/062,879
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/09/178,109
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2121
; TYPE: DNA
; ORGANISM: human
US-10-062-879-1
```

```
Query Match          24.3%; Score 34; DB 12; Length 2121;
Best Local Similarity 88.1%; Pred. No. 0.013;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATGTCATGAGCA 52
Db      2072 GCATGCTCCCGCGCCGATGCGCGGATTTTACAAGACA 2031
```

RESULT 7

```
US-09-924-400-175
; Sequence 175, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
```

```
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 175
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; OTHER INFORMATION: n = A,T,C or G
US-09-924-400-175
```

```
Query Match          24.1%; Score 33.8; DB 9; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.011;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATGTCATGAGCACTAACCCTTG 64
Db      25 GCATGNTCCCGCGCCGATGCGCGGATGCGGATGCGGATGCTCTCAAGGCTTG 78
```

RESULT 8

```
US-09-810-936-175
; Sequence 175, Application US/09810936
; Patent No. US20020068285A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Linda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.419C11
; CURRENT APPLICATION NUMBER: US/09/810,936
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-810-936-175
```

```
Query Match          24.1%; Score 33.8; DB 10; Length 484;
Best Local Similarity 75.9%; Pred. No. 0.011;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY      11 GCATGCTCCCGCGCCGATGCGCGGATGTCATGAGCACTAACCCTTG 64
Db      25 GCATGNTCCCGCGCCGATGCGCGGATGCGGATGCGGATGCTCTCAAGGCTTG 78
```

RESULT 9

US-09-429-755-175
; Sequence 175, Application US/09429755A
; Patent No. US2002011467A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; APPLICANT: Retter, Marc W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.419c6
; CURRENT APPLICATION NUMBER: US/09/429,755A
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 484
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(484)
; OTHER INFORMATION: n = A,T,C or G
US-09-429-755-175

Query Match
Best Local Similarity 24.1%; Score 33.8; DB 10; Length 484;
Matches 41; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 11 GCATGCTCCGGCGCATGGCGGCGGATTCATGAGCACTAAACCTTG 64
Db 25 GCATGNTCCGGCGCCGCGCATGGCGGCGGATTCGCGTTCCTCGAGCCTTG 78

RESULT 10

US-09-772-134B-80
; Sequence 80, Application US/09772134B
; Patent No. US20020144310A1
; GENERAL INFORMATION:
; APPLICANT: Southern Illinois University
; APPLICANT: Lightfoot, David
; APPLICANT: Messem, Khalid
; TITLE OF INVENTION: ISOLATED POLYNUCLEOTIDES AND POLYPEPTIDES RELATING TO LOCI UNDERL
; TITLE OF INVENTION: RESISTANCE TO SOYBEAN CYST NEMATODE AND SOYBEAN SUDDEN DEATH SYN
; FILE REFERENCE: 1268/4/2
; CURRENT APPLICATION NUMBER: US/09/772,134B
; CURRENT FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: 60/178,811
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 80
; LENGTH: 711
; TYPE: DNA
; ORGANISM: soybean
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(711)
; OTHER INFORMATION: n is an undetermined nucleotide (dntp, dctp, dgrp, or dttt)
US-09-772-134B-80

Query Match
Best Local Similarity 23.7%; Score 33.2; DB 10; Length 711;
Matches 35; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 11 GCATGCTCCGGCGCGCATGGCGGCGGATTCATGCA 48
|||||

Db 45 GCATGCTCCGGCGCCGCGCATGGCGGATTCGCTTAA 82

RESULT 11

US-10-112-267-52/c
; Sequence 52, Application US/10112267
; Publication No. US20030068678A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David A.
; APPLICANT: Cohen, Robert
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Lawrence, David A.
; APPLICANT: Levine, Arnold J.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: P1176R2
; CURRENT APPLICATION NUMBER: US/10/112,267
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/182,145B
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/063,704
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/073,612
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/081,695
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 52
; LENGTH: 163
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1-163
; OTHER INFORMATION: Sequence is synthesized
US-10-112-267-52

Query Match
Best Local Similarity 23.3%; Score 32.6; DB 9; Length 163;
Matches 35; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 11 GCATGCTCCGGCGCGCATGGCGGCGGATTCATGAG 49
Db 123 GCATGCTCCGGCGCCGCGCATGGCGGCGGATTAATCACTAG 85

RESULT 12

US-10-043-875-841/c
; Sequence 841, Application US/10043875
; Publication No. US20030054339A1
; GENERAL INFORMATION:
; APPLICANT: De Smelt, Koenraad
; APPLICANT: Stuyver, Lieven
; TITLE OF INVENTION: Method for detection of drug-induced mutations in the HIV reve
; TITLE OF INVENTION: Transcriptionase Gene
; FILE REFERENCE: 11362-0033-NPUS01 (INNS-033)
; CURRENT APPLICATION NUMBER: US/10/043,875
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 60/286,102
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: EP 01870085.6
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: EP 01870005.4
; PRIOR FILING DATE: 2001-01-11
; NUMBER OF SEQ ID NOS: 884
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 841
; LENGTH: 669
; TYPE: DNA

ORGANISM: Human immunodeficiency virus
US-10-043-875-841

Query Match 22.4%; Score 31.4; DB 9; Length 669;
Best Local Similarity 97.0%; Pred. No. 0.092;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
|||||
DB 626 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 594

RESULT 13

US-09-772-134B-82
Sequence 82, Application US/09772134B
Patent No. US20020144310A1
GENERAL INFORMATION:
APPLICANT: Southern Illinois University
APPLICANT: Lightfoot, David
APPLICANT: Meksem, Khalid
TITLE OF INVENTION: ISOLATED POLYNUCLEOTIDES AND POLYPEPTIDES RELATING TO LOCI UNDERL
TITLE OF INVENTION: RESISTANCE TO SOYBEAN CYST NEMATODE AND SOYBEAN SUDDEN DEATH SYM
FILE REFERENCE: 1268/4/2
CURRENT APPLICATION NUMBER: US/09/772,134B
CURRENT FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: 60/1178,811
PRIOR FILING DATE: 2000-01-28
NUMBER OF SEQ ID NOS: 122
SOFTWARE: PatentIn version 3.0
SEQ ID NO 82
LENGTH: 713
TYPE: DNA
ORGANISM: soybean
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(713)
OTHER INFORMATION: n is an undetermined nucleotide (dntp, dcrp, dgrp, or dtrp)
US-09-772-134B-82

Query Match 22.4%; Score 31.4; DB 10; Length 713;
Best Local Similarity 97.0%; Pred. No. 0.093;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
|||||
DB 46 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 78

RESULT 14

US-09-894-916-6/c
Sequence 6, Application US/09894916
Publication No. US20030032089A1
GENERAL INFORMATION:
APPLICANT: Hunter, Fiona F.
APPLICANT: Bidochka, Michael J.
TITLE OF INVENTION: Isolated Cocoon Silk Protein From Simulium vittatum And Nucleic A
FILE REFERENCE: 1468-001A
CURRENT APPLICATION NUMBER: US/09/894,916
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: 60/214992
PRIOR FILING DATE: 2000-06-29
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 831
TYPE: DNA
ORGANISM: Simulium vittatum
US-09-894-916-6

Query Match 22.4%; Score 31.4; DB 9; Length 831;
Best Local Similarity 97.0%; Pred. No. 0.096;

Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATTGT 43
|||||
DB 818 GCATGCTCCGGCCGCGCATGGCGGGGATTAT 786

RESULT 15

US-09-778-320-250
Sequence 250, Application US/09778320
Patent No. US20010034052A1
GENERAL INFORMATION:
APPLICANT: Dillon, David C.
APPLICANT: Day, Craig H.
APPLICANT: Jiang, Yugu
APPLICANT: Houghton, Raymond L.
APPLICANT: Mitcham, Jennifer
APPLICANT: Wang, TongTong
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.491C5
CURRENT APPLICATION NUMBER: US/09/778,320
CURRENT FILING DATE: 2001-02-06
NUMBER OF SEQ ID NOS: 301
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 250
LENGTH: 215
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(215)
OTHER INFORMATION: n=A,T,C or G
US-09-778-320-250

Query Match 22.1%; Score 31; DB 10; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCATGCTCCGGCCGCGCATGGCGGGGATT 41
|||||
DB 20 GCATGCTCCGGCCGCGCATGGCGGGGATT 50

Search completed: May 25, 2003, 20:21:58
Job time : 45.2464 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:52:25 ; Search time 487.844 Seconds
(without alignments)
4647.728 Million cell updates/sec

Title: US-09-743-533-14
Perfect score: 140
Sequence: 1 atagaatacagcatcgtccc.....gagctcgaattcgccctata 140

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_estro:*
9: em_estro:*
10: gb_estl2:*
11: gb_estl2:*
12: gb_estl3:*
13: gb_estl4:*
14: gb_estl5:*
15: em_estlun:*
16: em_estlun:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.4	27.4	521	9 AA255342	AA255342 T3598 Bio
2	32.8	23.4	402	9 AJ251073	AJ251073 AJ251073
3	32.4	23.1	686	13 BM579445	BM579445 170006872
4	32.4	23.1	733	13 B1273615	B1273615 A1241_1e
5	32.2	23.0	431	17 AO058015	AO058015 CIT-HSP-2
6	32.2	23.0	512	17 AZ366183	AZ366183 IM0115K24

C 7	32.2	23.0	559	13 B1262739	B1262739 EST003 Su
C 8	32.2	23.0	719	13 B1262738	B1262738 EST002 Su
C 9	32.2	23.0	729	13 B1262743	B1262743 EST008 Su
C 10	32.2	22.9	373	17 B30807	B30807 HS-1003-A2-
C 11	32.2	22.9	417	14 BQ094068	BQ094068 040802-27
C 12	32.2	22.9	442	17 AC015890	AC015890 CIT-HSP-2
C 13	31.8	22.7	237	14 BP018827	BP018827 BP018827
C 14	31.8	22.7	497	17 A1133472	A1133472 HA2082 Hu
C 15	31.8	22.7	1232	9 AB021948	AB021948 AB021948
C 16	31.6	22.6	362	17 B16177	B16177 347C21.TP C
C 17	31.4	22.4	446	17 B38957	B38957 HS-1048-B2-
C 18	31	22.1	150	9 AA012772	AA012772 RPU2008GC
C 19	31	22.1	185	9 A1738079	A1738079 ESTRGT02
C 20	31	22.1	187	9 AB029050	AB029050 AB029050
C 21	31	22.1	211	10 AM888331	AM888331 E45 Rat d
C 22	31	22.1	252	9 A1738086	A1738086 ESTRGT09
C 23	31	22.1	258	13 BM139628	BM139628 Gm-R2 Soy
C 24	31	22.1	277	9 AA052854	AA052854 T3513 Bio
C 25	31	22.1	278	13 B1203157	B1203157 PC89R Pea
C 26	31	22.1	319	9 AA585085	AA585085 T3655 Bio
C 27	31	22.1	330	9 AA585076	AA585076 T3645 Bio
C 28	31	22.1	335	9 AA585077	AA585077 T3646 Bio
C 29	31	22.1	365	9 AA052853	AA052853 T3507 Bio
C 30	31	22.1	393	12 BG006643	BG006643 RC2-GN026
C 31	31	22.1	408	13 B1203089	B1203089 PC89R1 Pe
C 32	31	22.1	438	10 AM888346	AM888346 MEG4 Rat
C 33	31	22.1	474	14 BQ094058	BQ094058 040802.17
C 34	31	22.1	581	13 BM139960	BM139960 Gm-R2 Soy
C 35	31	22.1	601	10 AM224940	AM224940 EST006 Pl
C 36	31	22.1	629	10 AV735727	AV735727 AV735727
C 37	31	22.1	4137	11 AF092132	AF092132 Homo sapi
C 38	30.8	22.0	190	9 AA052864	AA052864 T3541 Bio
C 39	30.8	22.0	377	9 AT006477	AT006477 AT006477
C 40	30.8	22.0	480	10 BEA11794	BEA11794 ISC009 ID0
C 41	30.6	21.9	468	13 BM359493	BM359493 GA_Ea002
C 42	30.6	21.9	565	9 A1728509	A1728509 BNLG1109
C 43	30.6	21.9	583	12 BG443394	BG443394 GA_Ea002
C 44	30.6	21.9	593	9 A1728966	A1728966 BNLG1121
C 45	30.6	21.9	639	17 AO110909	AO110909 CIT-HSP-2

ALIGNMENTS

RESULT 1
LOCUS AA255342 521 bp mRNA linear EST 14-MAR-1997
DEFINITION T3598 Bloodstream form of serodeme ILTat1.1 Trypanosoma brucei
brucei cDNA 5', mRNA sequence.
ACCESSION AA255342
VERSION AA255342.1 GI:1889945
KEYWORDS EST.
SOURCE Trypanosoma brucei brucei.
ORGANISM Trypanosoma brucei brucei.
Eukaryota: Euklenozoa: Kinetoplastida: Trypanosomatidae:

REFERENCE 1 (bases 1 to 521)
Osanya,A., Pelle,R. and Murphy,N.B.
Identification and characterisation of differentially expressed
genes of Trypanosoma brucei brucei
Unpublished (1997)
JOURNAL
COMMENT
Contact: Osanya A
Unit 3
International Livestock Research Institute
Box 30709, Nairobi, Kenya
Tel: 254 2 630 743
Fax: 254 2 631 499
Email: a.osanya@cgnel.com
Seq primer: M13 primer.
Location/Qualifiers
1..521
/organism="Trypanosoma brucei brucei"
/strain="Clone A4"

FEATURES

Source

/db_xref="taxon:5702"
/clone_lib="Bloodstream form of serodeme ITat1.1"
/note="CDNs were generated from poly (A+) enriched mRNA prepared from different developmental stages of T.b. brucei by reverse transcription followed by PCR amplification using mini-exon and oligo(dT) primers. The cDNA generated were utilized in RABES-PCR coupled with differential hybridisations to identify differentially expressed mRNA transcripts. The products which showed to be differentially expressed were cloned pGEM-T vector. These differentially expressed mRNA transcripts were (are being) sequenced to generate differentially expressed sequence tags."

BASE COUNT 151 a 140 c 117 g 113 t
ORIGIN

Query Match 27.4%; Score 38.4; DB 9; Length 521;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 60; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCGCCGATGTCATGAGCAGCACTAACCCCTGGCAGCG 70
|||||
Db 142 GCATGCTCCCGCGCCGATGTCGCGCGGATTCGGCGCCAGCACCATTATTCGCACT 201
|||||

OY 71 CCCCCACAGCTTCATGATGAGTATGCTCTA 106
|||||
Db 202 TCCCTCATGCACTCAAGCACAGATGTTAGTGGGATA 237
|||||

RESULT 2
AJ251073 402 bp mRNA linear EST 29-NOV-1999
LOCUS AJ251073 Rattus norvegicus library (Ma MZ) Rattus norvegicus cDNA
DEFINITION clone #1, mRNA sequence.
ACCESSION AJ251073
VERSION AJ251073.1 GI:6470363
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 402)
Ma,Z.M., Yang,Y., Xie,Y.Z. and Fan,M.
The DB-PCR studies of the candidate genes of rat brain under system hypoxia

JOURNAL Unpublished (1999)
COMMENT Contact: Ma MZ
Department of Neurobiology Sciences
Institute of Basic Medical Sciences
No.27, Tai Pin Road, Beijing, 100850, China.

FEATURES
Source
1..402
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="Rattus norvegicus library (Ma MZ)"
/note="Organ: brain; Vector: pGEM-T; with treatment by endostatin, the differential expressed CDNs were cloned into the pGEM-T vector. The 3' anchor primer is AAGCTTTTCTTTT, and the 5' arbitrary primer is 10-mer oligo. Individual clone was sequenced with T7 promoter primer."

BASE COUNT 113 a 89 c 87 g 112 t 1 others
ORIGIN

Query Match 23.4%; Score 32.8; DB 9; Length 402;
Best Local Similarity 94.4%; Pred. No. 16;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 11 GCATGCTCCCGCGCCGATGTCATGATTCAT 46
|||||
Db 58 GCATGCTCCCGCGCCGATGTCATGATTCAT 93
|||||

RESULT 3
BM579445 686 bp mRNA linear EST 22-FEB-2002
LOCUS BM579445
DEFINITION 17000687238519 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
19600449682280 5', mRNA sequence.
ACCESSION BM579445
VERSION BM579445
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 686)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C., and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AZX row: G column: 02
Seq primer: M13 Reverse.

FEATURES
source
1..686
/organism="Anopheles gambiae"
/strain="RSP-St (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449682280"
/clone_lib="A.Gam.ad.cDNA.blood1"
/dev_stage="Adult"
/lab_host="DHI0b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"

BASE COUNT 141 a 157 c 199 g 189 t
ORIGIN

Query Match 23.1%; Score 32.4; DB 13; Length 686;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 12 CATGCTCCCGCGCCGATGTCATGAGCAGCACTAACCCCTGGCAGCGTC 71
|||||
Db 281 CAGGCGCTCGCGCGCTTGGCCGCGCTGCTTCATTTCGACGAGAAACCGGCAAGAAC 340
|||||

OY 72 CCCCCACAGCTTCATGATGAGTATGAGTATG 101
|||||
Db 341 GCCACACAGTTAATGATGTTGATATG 370
|||||

RESULT 4
BI273615 733 bp mRNA linear EST 12-JUN-2002
LOCUS BI273615 Aplysia metacerebral cell cDNA library Aplysia californica
DEFINITION cDNA 5', mRNA sequence.
ACCESSION BI273615
VERSION BI273615.1 GI:18020448
KEYWORDS EST.
SOURCE California sea hare.
ORGANISM Aplysia californica
Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia;
Anaspeidae; Aplysiidae; Aplysia.
1 (bases 1 to 733)

449 Chem/Biochem, Manhattan, KS 66506, USA
Tel: 785 532 6964
Fax: 785 532 7278
Email: kanostek@ksu.edu
Seq primer: SP6

High quality sequence stop: 730.

FEATURES
source
Location/Qualifiers
1..729
/organism="Manduca sexta"
/db_xref="taxon:7130"
/clone="008"
/clone_lib="Subtracted larval fat body library,
bacteria-induced genes"
/tissue_type="Fat Body"
/dev_stage="5th instar larva"
/note="Vector: pGEM-T Vector; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 158 a 215 c 187 g 165 t 4 others

Query Match 23.0%; Score 32.2; DB 13; Length 729;
Best Local Similarity 91.9%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGCGGATTCATG 47
141 GCATGCTCCGCGCCGATGCGCGGATTCATG 105

RESULT 10

B30807 373 bp DNA linear GSS 17-OCT-1997
LOCUS HS-1003-A2-B01-MR.abi CIT Human Genomic Sperm Library C Homo
DEFINITION sapiens genomic clone Plate=CT 497 Col=2 Row=C, DNA sequence.

ACCESSION B30807
VERSION B30807.1 GI:2530176

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 373)
Mahairas G.G., Zackrone K.D., Smith T., Tipton S., Schmidt S.,
Tralcoff R., Abdjian C., Blanchard A., West A. and Hood L.E.
Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
Unpublished (1997)

JOURNAL
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 497 row: C column: 2
Class: BAC ends
High quality sequence stop: 373.

FEATURES
source
Location/Qualifiers
1..373
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 497 Col=2 Row=C"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 104 a 110 c 80 g 76 t 3 others

Query Match 22.9%; Score 32; DB 17; Length 373;
Best Local Similarity 79.2%; Pred. No. 27;
Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 93 TGCAGTATGCTAGAGATCCGGTACGAGCTCGAATTCACCCATA 140
16 TGCAGTATGCTAGAGATCCGGTACGAGCTCGAATTCACCCATA 63

RESULT 11
B0094068 417 bp mRNA linear EST 08-APR-2002
LOCUS 040802_27 Rat male liver Rattus norvegicus CDNA 5', mRNA sequence.
DEFINITION B0094068
ACCESSION B0094068.1 GI:20075285
KEYWORDS EST.
SOURCE Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 417)

REFERENCE
AUTHORS Lee Y. and Xu C.
TITLE Expressed sequence tags from cDNA subtractive library following
short interval successive partial hepatectomy in rat liver
regeneration
Unpublished (2002)

JOURNAL
COMMENT Contact: Lee Yuchang
Henan Bioengineering Key Lab
Henan Normal Univ
Jianshe Road, Xinxiang City, P.R.China
Tel: 0086373328084
Fax: 0086373326524
Email: yuchang910@163.com
Seq primer: M13 Forward.
Location/Qualifiers
1..417
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="Rat male liver"
/sex="male"
/tissue_type="liver"
/note="Vector: pT73D-Pac (Pharmacia)-with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UTR-D01
library is a normalized Rat Cell Line R3327-5A library
(m3A) constructed in pT73D PAC vector according to the
procedure described by Bonaldo, Lennon & Soares
(Normalizaton and Subtraction: Two Approaches to
Facilitate Gene Discovery. Genome Research 6: 791-806,
1996). The oligonucleotide used to prime first strand
synthesis contained the sequence tag GGACATGATC between
the Not I cloning site and G118 stretch. The Rat Cell Line
R3327-5A was provided by Mary Hendrix of the University of
Iowa."

FEATURES
source

BASE COUNT 66 a 124 c 97 g 130 t

Query Match 22.9%; Score 32; DB 14; Length 417;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GCATGCTCCGCGCCGATGCGCGGATTC 42
19 GCATGCTCCGCGCCGATGCGCGGATTC 50

RESULT 12
A0015890 442 bp DNA linear GSS 09-JUN-1998
LOCUS CIT-HSP-2314E8.TF CIT-HSP Homo sapiens genomic clone 2314E8, DNA
DEFINITION sequence.
ACCESSION A0015890
VERSION A0015890.1 GI:3194626
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

							/db xref="taxon:7719"										
							/clone="cl1y05c06"										
							/clone_lib="Nori Satoh unpublished cDNA library, larva"										
							/tissue_type="whole animal"										
							/dev_stage="larva"										
							/note="Vector: pBluescript SK"										
BASE COUNT	57 a	56 c	65 g	58 t	1	others											
ORIGIN																	
Query Match	22.7%	Score 31.8:	DB 14:	Length 237:													
Best Local Similarity	83.7%	Pred. No. 25:	Matches 7:	Indels 0:	Gaps 0:												
Matches	36:	Conservative	0:	Mismatches	7:	Indels	0:	Gaps									
Oy	11 GCATGCTCCCGGCCCATGGCCGCGSGATTGTTCATGAGCAAC 53																
Db	184 GCATGCTCCCGGCCCATGGCCGCGSGATTATATCAAGATAA 142																
RESULT 14	A1133472	497 bp	mRNA	linear	EST 11-NOV-1999												
LOCUS	A1133472/c																
DEFINITION	HA2082 Human fetal liver cDNA library Homo sapiens CDNA, mRNA																
ACCESSION	A1133472.1	GI:6360788															
VERSION	A1133472.1	GI:6360788															
KEYWORDS	EST.																
SOURCE	human.																
ORGANISM	Homo sapiens																
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.																
TITLE	1 (bases 1 to 497)																
JOURNAL	Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.																
COMMENT	and He,F. Expression profile analysis of a human fetal liver CDNA library Unpublished (1998) Contact: Yongtao Yu Department of Hematology Beijing Institute of Radiation Medicine 27 Taiping Road Beijing 100850, P.R.China Tel: 0086-10-68159479 Fax: 0086-10-68214653 Email: yyl48@yahoo.com.																
FEATURES	location/Qualifiers																
source	1..497																
	/organism="Homo sapiens"																
	/db_xref="taxon:9606"																
	/clone_lib="Human fetal liver CDNA library"																
	/tissue_type="liver"																
	/dev_stage="fetal"																
	/lab_host="MC1061/PJ3"																
	/note="Vector: pCDNA1"																
BASE COUNT	128 a	115 c	141 g	113 t													
ORIGIN																	
Query Match	22.7%	Score 31.8:	DB 9:	Length 497:													
Best Local Similarity	76.5%	Pred. No. 35:	Matches 12:	Indels 0:	Gaps 0:												
Matches	39:	Conservative	0:	Mismatches	12:	Indels	0:	Gaps									
Oy	11 GCATGCTCCCGGCCCATGGCCGCGSGATTGTATGAGCAACTAAACC 61																
Db	458 GCATGCTCCCGGCCCATGGCCGCGSGATTGAATGAGTGCTCAGTCC 408																
RESULT 15	AB021948	1222 bp	mRNA	linear	EST 07-JAN-1999												
LOCUS	AB021948																
DEFINITION	AB021948 Oryza sativa ms-bo-Taichung 65 immature anther Oryza																
ACCESSION	sativa cDNA clone DD2R3-1, mRNA sequence.																
VERSION	AB021948																

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

AUTHORS

1 (bases 1 to 1222)

TITLE

EST of immature anther of rice

JOURNAL

Unpublished (1999)

COMMENT

Contact: Toriyama K
Faculty of Agriculture, Laboratory of Plant Breeding
Tohoku University
Tatsunomidori-Amamiyamachi 1-1, Aobaku, Sendai, Miyagi 981-8555,
Japan
Email: torikin@bios.tohoku.ac.jp.
Location/Qualifiers

FEATURES

source

1. 1222
/organism="Oryza sativa"
/strain="ms-bo-Taichung 65"
/db_xref="taxon:4530"
/clone="DD2R3-1"
/clone_id="Oryza sativa ms-bo-Taichung 65 immature
anther"
/issue_type="anther"
/dev_stage="immature anther"
BASE COUNT 297 a 298 c 291 g 335 t 1 others
ORIGIN

Query Match

22.7%: Score 31.8; DB 9; Length 1222;

Best Local Similarity 94.3%; Pred. No. 52;

Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy

11 GCATGCTCCGCGCCCATGGCGCGGATTGTCA 45
|||||

DB 16 GCATGCTCCGCGCCCATGGCGCGGATTATTA 50

Search completed: May 25, 2003, 20:17:10
Job time : 491.844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:47:50 ; Search time 1766.52 Seconds

(without alignments)
7858.394 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 477

Sequence: 1 atgagcactaaccttg.....gcata tggagtatgtctag 477

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl : *

1:	gb_da:
2:	gb_hfg:
3:	gb_in:
4:	gb_om:
5:	gb_ov:
6:	gb_pat:
7:	gb_ph:
8:	gb_pl:
9:	gb_pr:
10:	gb_ro:
11:	gb_sts:
12:	gb_sy:
13:	gb_un:
14:	gb_vj:
15:	em_ba:
16:	em_fun:
17:	em_hum:
18:	em_in:
19:	em_nm:
20:	em_om:
21:	em_or:
22:	em_ov:
23:	em_pat:
24:	em_pl:
25:	em_ph:
26:	em_ro:
27:	em_sts:
28:	em_un:
29:	em_vj:
30:	em_hfg_hum:
31:	em_hfg_in:
32:	em_hfg_other:
33:	em_hfg_mus:
34:	em_hfg_pln:
35:	em_hfg_rod:
36:	em_hfg_nam:
37:	em_hfg_vrl:
38:	em_hsg:
39:	em_hsgo_hum:
40:	em_hsgo_mus:
41:	em_hsgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	378.8	79.4	1420	8	HYHORI17	X60037 H.vulgare h
2	329.6	69.11	2806	8	S66938	S66938 C-hordein [
3	331.4	48.35	3325	8	AF280606	AF280606 Trilicium
4	230.6	48.3	9248	8	SCAF000227	AF000227 Secale ce
5	229	48.0	1832	8	SCSECB	X60295 S.cereale S
6	224.2	47.0	1342	8	SCSECB	X60294 S.cereale S
7	219.6	45.4	3789	8	AF280605	AF280605 Trilicium
8	216.6	45.4	2065	8	BLYHORDCA	M36941 Hordeum vul
9	200.6	42.1	1858	8	AB059812	AB059812 Trilicium
10	148.2	31.1	4971	8	AF201084	AF201084 Secale ce
11	140.4	29.4	927	8	AF175312	AF175312 Trilicium
12	140.4	29.4	1178	8	AF234650	AF234650 Trilicium
13	140.4	29.4	1397	8	WHTGLGCV	M16064 Wheat (T..ae
14	139.4	29.2	771	8	AF234642	AF234642 Trilicium
15	139.4	29.2	1227	8	AF234644	AF234644 Trilicium
16	135.2	28.3	5718	8	AF234647	AF234647 Trilicium
17	134.6	28.2	730	8	AS0389659	AJ389659 Aegilops
18	134.6	28.2	757	8	TAE38666	AJ389666 Trilicium
19	134.6	28.2	777	8	TAE38668	AJ389668 Trilicium
20	134.6	28.2	778	8	TSP389674	AJ389674 Trilicium
21	134.6	28.2	792	8	TC0389669	AJ389669 Trilicium
22	134.6	28.2	798	8	WHTGLGCV	M16060 Wheat (T..ae
23	134.6	28.2	800	8	TMA389671	AJ389671 Trilicium
24	134.6	28.2	801	8	TMA389673	AJ389673 Trilicium
25	134.6	28.2	801	8	TSP389676	AJ389676 Trilicium
26	134.6	28.2	809	8	TMA389672	AJ389672 Trilicium
27	134.6	28.2	850	8	TC0389670	AJ389670 Trilicium
28	134.6	28.2	947	8	AF120267	AF120267 Trilicium
29	134.6	28.2	956	8	AF144104	AF144104 Trilicium
30	133.6	28.0	2086	8	WHTGGLN	M36999 Wheat gene
31	132	27.7	842	8	TDBGCL	X53412 Wheat gene
32	131.8	27.6	750	8	TAE389667	AJ389667 Trilicium
33	131.8	27.6	762	8	TVA389678	AJ389678 Trilicium
34	131.4	27.5	731	8	AS0389680	AJ389680 Aegilops
35	130.6	27.4	731	8	TSP389675	AJ389675 Trilicium
36	129.6	27.2	746	8	TVA389677	AJ389677 Trilicium
37	128.2	26.9	6443	8	AF234649	AF234649 Trilicium
38	127.4	26.7	604	8	AF2389685	AJ389685 Aegilops
39	126.6	26.5	156533	2	AC117070	AC117070 Dictyoste
40	125.8	26.4	622	8	AS0389687	AJ389687 Aegilops
41	124.6	26.1	97683	2	AC116548	AC116548 Dictyoste
42	124.4	26.1	718	8	TDU389703	AJ389703 Trilicium
43	124.4	26.1	886	8	TDU389704	AJ389704 Trilicium
44	124.2	26.0	607	8	AS0389688	AJ389688 Aegilops
45	124.2	26.0	611	8	AS0389684	AJ389684 Aegilops

ALIGNMENTS

RESULT 1				
HVHOR117				
LOCUS	HVHOR117	1420 bp	DNA	linear
DEFINITION	H. vulgare hor1-17 gene for C-hordein.			
ACCESSION	X60037			
VERSION	X60037.1	GI:19000		
KEYWORDS	endosperm; seed storage protein.			
SOURCE	Hordeum vulgare subsp. vulgare.			
ORGANISM	Hordeum vulgare subsp. vulgare			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
	Poideae; Triticeae; Hordeum.			
REFERENCE	1 (bases 1 to 1420)			
AUTHORS	Entwistle,J.			
TITLE	Direct Submission			

	FEATURES	
	source	Location/Qualifiers 1..1420
		/organism="Horddeum vulgare subsp. vulgare"
		/cultivar="Bomi"
		/db_xref="taxon:112509"
		/clone="-horl-17"
		109..136
	promoter	/note="-300 element"
		229..250
	repeat_region	/note="caca box"
		327..334
	TATA_signal	432..1214
	CDS	/note="unnamed protein product"
		/codon_start=1
		/protein_id="CA4A2642.1"
		/db_xref="GI:19001"
		/db_xref="SPIRMBL:Q40053"
		/translation="MKTFFFLVLIAMWSIVTARQLNFSSELASPDQSQTQQPYPQ NYLPQKFPVPQSQFHTPPQYFYLPBELPPOVITPLPQPQPFPOQSPQGLPBPQC OPWQPPQFPQPEIPQDPQGFQPPQCPQDPPQOIIIFQDPQASYPQDPQPF QDPQVPQDPQQAAPLOPQQAAPLOPQPFQSGEOLIQDPFPLAQNPBPQDPQ PLPQDPQFROGAELIIPQDPQPLPQNPQPTGQITMSW"
		1302..1307
polyA_signal		1356..1361
polyA_signal		444 C
BASE COUNT	518 A	176 G
ORIGIN		282 T

Query Match	Similarity	Score	378-8	DB	Length
Best Local	Conservative	92.2%	0	Mismatches	32
Matches	411	Indels	3	Gaps	1
QY	4	AGGCAACTAAACCCCTGGACGCAAGAGTTGGCAATCACCACACAAATCATATCTGCGCGAC	63		
Db	492	AGGCAACTAAACCCCTAGACGCAAGAGTTGGCAATCACCACAAATCATATCTGCGCGAG	551		
QY	64	CCATATCCCAAAACCCCATCTACCGCAAAAACATTTCCTCATGTGACAGAACCGTTTCAC	123		
Db	552	CCATATCCCAAAACCCCATCTACCGCAAAAACATTTCCTCATGTGACAGAACCGTTTCAC	611		
QY	124	ACACCCGAACAATATTTCCCTATCTACAGAGAGATTTGTTCCCAATATCAATATACA	183		
Db	612	ACACCCGAACAATATTTCCCTATCTACAGAGAGATTTGTTCCCAATATCAATATACA	671		
QY	184	ACCCCCCTGACACGACAACCATTTCCCCCAACACACAGAACCCCTCTCTCGGGCC	243		
Db	672	ACCCCCCTGACACGACAACCATTTCCCCCAACACACAGAACCCCTCTCTCGGGCC	731		
QY	244	CACACAACCTTTCCCTGGGCAACGACAAACACATTTCCCGACGCCCAAGAACCAATTCG	303		
Db	732	CACACAACCTTTCCCTGGGCAACGACAAACACATTTCCCGACGCCCAAGAACCAATTCG	791		
QY	304	CACACAACCTTTCCCTGGGCAACGACAAACATTTCCCGACGCCCAAGAACCAATTCGA	363		
Db	792	CACACAACGACAAACCAATTTCCCGACGAGAACGACAAACCAATTTCCCGACGACCAACA	851		
QY	364	CAATTAATTTTCAGCAACCCCAACATCATACCTGTGCAACCTGACAGCCATTTCT	423		

Db	QY	RESULT 2	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM
852	CAAAATATTTTCAGACCAACCCCAATCTATCCCTGTGCAACCTCAGACGCAATTCT	911						
424	CAACAACCTCAACACGATCCCCCAACA	449						
912	---CAACCTCAACGAGTCCCCACCA	934						
S66938		2806 bp	DNA	linear	PLN 19-JAN-1994			
C-hordein								
[Hordeum vulgare=barley, M564, Genomic, 2806 nt].								
S66938		1	GI:442523					
Hordeum vulgare								
Hordeum vulgare M564.								
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;								
Pooidae; Triticeae; Hordeum.								
1 (bases 1 to 2806)								
Saitova,O.V., Mekhedov,S.L., Zhelnin,L.G., Khokhlova,T.A. and								
Anan'ev,E.V.								
Nucleotide sequence of the barley C-hordein gene								
Genetika 29 (7), 1070-1079 (1993)								
93380629								
8396543								
Genbank staff at the National Library of Medicine created this								
entry (NCBI g1bbseq 139925) from the original journal article.								
This sequence comes from Fig. 2.								

Location/Qualifiers
1. .2806

REFERENCE	1 (bases 1 to 2806)
AUTHORS	Sainova,O.V., Mekhedov,S.L., Zhelnin,L.G., Khokhlova,T.A. and Anan'ev,E.V.
TITLE	Nucleotide sequence of the barley C-hordein gene
JOURNAL	Genetika 29 (7), 1070-1079 (1993)
MEDLINE	93380629
PubMed	8396543
REMARK	GenBank staff at the National Library of Medicine created this

GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 139925] from the original journal article. This sequence comes from Fig. 2.

```

/organism="Hordium vulgare"
/db_xref="taxon:4513"
1084. 2016
/gene="C-hordein"
1084. 2016
/gene="C-hordein"
/note="this sequence comes from Fig. 2"
/codon_start=1
/product="C-hordein"
/protein_id="AAP28161.1"
/db_xref="GI:142524"
/translation="MKTFTLVLLAMWMSIVTARGLNPSSELOSPOQSYLOQPYPO
NLTLEQFPVYQDPFTFTFQDLPFLPEBELSPQIQLPFLQDPNPPQDPDLPDPQ
PFWHQDPQFPQDPPLPQDPQFPNQDPQIISQDPQDFQDPQDPQDPQDPFW
QDQDFPDPQDPFLQDPQDPFWQDPQDPFQDPQDPFAHQDPQDFSQDPQDPFLQ
DQDFPQDPQDPFPQDPQDPIIFQDPQDSYFVQDPQDPQDPVQDPQDQASPLQDPQ
PFGSGQLLPQDPQDPFLQDPQDPYQDQITWSV"
BASE COUNT      958 a      711 c      420 g      714 t      3 others
ORIGIN

```

	Query Match	69.1%	Score 329.6	DB 8	Length 2806
	Best Local Similarity	85.1%	Pred. No. 2.6e-56		
	Matches 400	Conservative 0	Mismatches 49	Indels 21	Gaps 2
Oy	4	AGGCAACTAACCCCTTCGACGCCAAGAGTGTGCATCAACCAACATCATATATCTGGCCAG	63		
Db	1144	AGGCAACTAACCCCTTCGACGCCAAGAGTGTGCATCAACCAACATCATATCTGGCCAG	1203		
Oy	64	CGATATCCCAAAACCCATATCTACGGCAAAACCATTTGCATGAGCAACCGTTTAC	123		
Db	1204	CGATATCCCAAAACCCATATCTACCGCAACCAATTTGCATGAGCAACCGTTTAC	1263		
Oy	124	ACACGCCAACATATTTCCCTATCTACAGAGGAATTTGCCCAATATCAATATACCA	183		
Db	1264	ACACGCCAACATATTTCCCTATCTACAGAGGAATTTGCCCAATATCAATATACCA	1323		
Oy	184	ACCCGCCCTFACAGCAGCAACCAATGCCCAAGACAGCAACCAACCTTCTCGGGCC	243		
Db	1324	ACCCGCCCTFACAGCAGCAACCAATGCCCAAGACAGCAACCAACCTTCTCGGGCC	1383		
Oy	244	CAGCAACCAATTCCTCGCAACCAACCAACATTTTCCCGAGGCCCAAGAACCAATTC	303		

```

Db      1384 CAAACAACATTCCTCCGCGACCAACAACATTTCCCGAGCCCAACAACCAATTTCC 1443
QY      304 -----CAACACCATTTCCCTGGCAACGACAAACATTTCCCGAGCCCAAGAA 354
Db      1444 TACCAACCAACAAACCATTTCAACGACGACAAACAATATATATCCAGCAACCA 1503
QY      355 -----CCAAATTCAGCAAAATATTTTCAGACCAACCCCAACATCATACCTGTG 402
Db      1504 CAACCATTTCCCAACCAACCAACCACTTTTCCTACAGCCCAACCAATTTCCCTGTG 1563
QY      403 CAACCTCAACGACATTTCTCTCAACACCTCAACCAAGTCCCAACCAAC 452
Db      1564 CAACCAACAACCAATTTCCAGGCTCAACAACCAATTTCCCTGCAACC 1613

RESULT 3
AF280606 Locus AF280606 3925 bp DNA Linear PLN 02-OCT-2000
DEFINITION Triticum aestivum omega gliadin pseudogene, complete sequence.
ACCESSION AF280606
VERSION AF280606.1 GI:10444085
KEYWORDS Triticum aestivum.
SOURCE Triticum aestivum.
ORGANISM Triticum aestivum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 3925)
AUTHORS Hsia,C.C. and Anderson,O.D.
TITLE Isolation and characterization of wheat omega gliadin genes
JOURNAL Unpublished
2 (bases 1 to 3925)
AUTHORS Hsia,C.C. and Anderson,O.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Western Regional Research Center,
Agricultural Research Service, 800 Buchanan Street, Albany, CA
94710, USA

FEATURES
source location/Qualifiers
1..3925
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/chromosome="1"
/clone="omega3"
/dev_stage="embryo"
2372..2377
2477..3550
TATA_signal
CDS
TATA_signal
2477..3550
/feature="prolamine storage protein"
pseudo
/codon_start=1
/product="omega gliadin"
3026..3028
/feature="in-frame stop codon"
polya_signal
3626..3631
3683..3688
BASE COUNT 1327 a 995 c 559 g 1044 t
ORIGIN

Query Match 48.5%; Score 231.4; DB 8; Length 3925;
Best Local Similarity 74.5%; Pred. No. 1.2e-36;
Matches 327; Conservative 0; Mismatches 91; Indels 21; Gaps 2;

```

```

QY      220 CCACAACAACCTCTTCTCCGCCCCCAACAACCATTTCCCTGCGAACAACAACCAATTT 279
Db      2774 CCACAACAACCTTTTCCAGACCCCAACAACCATTTCTTGGCAACAACAACCAATTT 2833
QY      280 CCCCAGCCCCAAGAAGCAAT-----TCCCAACAACCATTTCCCTGGCAACGACAA 330
Db      2834 CCCCAGACCCCAACAATGCTTCCCTCTGTGCAACGACGACCATTTCCCGACGACCCAA 2893
QY      331 CAACCATTTCCCGAGCCCAAGAACCAATTC-----AACAAATATTTTCCAG 378
Db      2894 CAACCATTTCCCGAGCCCAACTACATTTCCCAACAACGACGACCAACAATATTTCCGAG 2953
QY      379 CAACCCCAACAATATACCTGTGCAACGCTCAACGACATTTCTCTCAACAACCTCAACA 438
Db      2954 CCAACCAACAACCATTTCCCTCTGTGCAACGACGACGACCATTTCCCGCAACCAATTTCCC 3013
QY      439 GTCCCCCAACAAGCTTCAT 457
Db      3014 CAGCAACCCCAATTAACCAT 3032

RESULT 4
SCAF000227 Locus SCAF000227 9248 bp DNA Linear PLN 31-MAY-1997
DEFINITION Secale cereale omega secalin gene, complete cds.
ACCESSION AF000227
VERSION AF000227.1 GI:2145024
KEYWORDS Secale cereale.
SOURCE Secale cereale.
ORGANISM Secale cereale.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Secale.
REFERENCE 1 (bases 1 to 9248)
AUTHORS Clarke,B.C., Mukai,Y. and Appels,R.
TITLE The Sec-1 locus on the short arm of chromosome 1R of rye (Secale
cereale)
JOURNAL Chromosoma 105 (5), 269-275 (1996)
MEDLINE 97094352
PUBMED 8939819
REFERENCE 2 (bases 1 to 9248)
AUTHORS Clarke,B. and Appels,R.
TITLE Sequence variation at the Sec-1 locus of rye
JOURNAL Plant Syst. Evol. (1997) In press
REFERENCE 3 (bases 1 to 9248)
AUTHORS Clarke,B. and Appels,R.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1997) Plant Industry, C.S.I.R.O., Clunies Ross,
Canberra, ACT 2601, Australia

FEATURES
source location/Qualifiers
1..9248
/organism="Secale cereale"
/db_xref="taxon:4550"
/chromosome="1"
/map="1RS"
/tissue_type="endosperm"
1..9248
/feature="there are 15 omega secalin genes at the Sec-1
locus; this sequence covers one complete unit of this
multigene family"
<2751..>3825
/product="omega secalin"
/feature="there are 15 omega secalin genes at the Sec-1
locus"
2752..3825
/function="seed storage protein"
/codon_start=1
/product="omega secalin"
/protein_id="AB58403.1"
/db_xref="GI:2145025"
/translation="MKTFILFVLAHTMSIITTAQUNPSDEGLSPQDPVPEKESYPO

```


QY	184	ACCCGCCGTAACAACAACAACGATTCCGCCAACAACCAACAACCTTTCTCGGCC	243
Db	364	ACCCGCATATTCACACCAACAACGATTCCGCCAGGCCAACAACCTTTCCCGAGCC	423
QY	244	CACAACAACATTTCCCTGGCAACCAACAACGATTTTCCGAGCCCAAGAACCATTCC	303
Db	424	CACAACAATTTTACCTTGCACACCAACAACAATTCATTTCCGAGCCCAACATTCAC	483
QY	304	CACAACA-----TTCCCGTGGCAACCAACAACGATTTTCCCGAGCCCAAGA	354
Db	484	CAGCAACCAACAACATTCGTTCCCGCAACACCCGAGAGACCAAGAACAATTTCCCGC	543
QY	355	CCAATTCACAATAATATTTTCAGCAACCCCAACAATATCCCTGTGCACCTCAACG	414
Db	544	CACCAACAACAATATTTTCCCGACCAACAACAACAACGATTTTCCCTGTCACTCAACA	603
QY	415	CGATTTCTCAACAACCTTCA	435
Db	604	CGATTTCCACAACAACGACAA	624

```

FEATURES
    source
        Location/Qualifiers
            1..3789
                /organism="Triticum aestivum"
                /db_xref="taxon:4565"
                /chromosome="1"
                /clone="omegaF20b"
                /dev_stage="embryo"
                2359..2364
                <2464..23441
                /product="omega gliadin storage protein"
                2464..3306
                /codon_start=1
    ORGANISM
        Triticum aestivum.
        Triticum aestivum
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
        Poideae; Triticeae; Triticum.
        1 (bases 1 to 3789)
        Hs1a,C.C. and Anderson,O.D.
        Isolation and characterization of wheat omega gliadin genes
        Unpublished
        2 (bases 1 to 3789)
        Hs1a,C.C. and Anderson,O.D.
        Direct Submission
        Submitted (20-JUN-2000) Western Regional Research Center,
        Agricultural Research Service, 800 Buchanan Street, Albany, CA
        94710, USA
    REFERENCE
        AUTHORS
        TITLE
        JOURNAL
        REFERENCE
        JOURNAL

```

[illegible]

[illegible]

BASE COUNT	1549.1554	258 g	466 t
ORIGIN			
Query Match	45.48;	Score 216.6;	DB 8; Length 2065;
Best Local Similarity	72.3%;	Pred. No. 1,1e-33;	
Matches 336;	Conservative 0;	Mismatches 99;	Indels 30; Gaps 3;
QY	4	AGGCAACTAAACCTTTCGAGCCGCAAGATTCGCAATCAGCAACAAATGATATCTGCGCCAG	63
Db	489	AGGCAAGCTAAACCTTACGACAGAGATGTGCAATACCAACAAACCAATTTCTGTAAACAA	548
QY	64	CCATATTCACAAAACCCATATCTCTCCGCAAAAACCAATTTCCATGAGCAACCGTTTCAC	123
Db	549	CAATCATATATCTGCACAACCAATATTCACCAACCAACCAATATCTACCGCAGCAACGTTCCCG	608
QY	124	ACACCCCAACAAATATTTTCCCTATCTACGAGAGGAAATGTTTCCCAATATCAAAATACCA	183
Db	609	ACACCCCAACAAATTTTCCCTATCTACGAGAGGAAATGTTTCCCAATATCAAAATACCA	668
QY	184	ACCCCTCTACAGCAACAACCAATTC-----CCCAACAACCAACAACCAACCTCTT	234
Db	669	AACCCCTCTACAGCAACAACCAATTC-----CCCAACAACCAACAACCAACCAACCTTTT	728
QY	235	CCCTGGGCCCCAAGCAACCAATTCCTGGCAACCAACAACCAATTTCCCGACCCCAAGAA	294
Db	729	CCTAGGCCCAACAACCAAAATCCCGAGCAACCAACAACCACTTTCCCGGCAACCAACAA	788
QY	295	CCAAT-----TCCCAACAACCAATTCCTGGCAACCAACAACCAACCAATTTCC	342
Db	789	CAAAATAGTACCCCAAGCAACCAACAACCAATTTCCCTGGCAACCAACAACCAACCTTTTCT	848
QY	343	CAGCCCAACAACCAAT-----TCACAATTAATTTTCGAGCAACCCCAACCAATCA	393
Db	849	CAGCCCAACAACCAATTTCTTTGGCAACCAACAACCAACCAATTTTCACCCCTTCAACACTA	908
QY	394	TACCTGTGCAACGTCACAGCAACCAATTTCCCAACCACTCAACA	438
Db	909	TAGCCCTGTGCAAGCAACAACCAATTTCCCTGTGCAACCTCAACTA	953

```

RESULT 9
AB059812 LOCUS 1858 bp DNA linear PLN 03-AUG-2002
DEFINITION Triticum aestivum pseudogene for omega gliadin.
ACCESSION AB059812
VERSION AB059812.1 GI:16904641
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (cultivar:Chinese Spring) DNA.
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1
Masoudi-Nejad, A., Nasuda, S., Kawabe, A. and Endo, T. R.
Molecular cloning, sequencing, and chromosome mapping of a
1A-encoded Omega-type Prolamin sequence from Wheat
Genome 45, 661-669 (2002)
2 (bases 1 to 1858)
Masoudi-Nejad, A. and Nasuda, S.
Direct Submission
Submitted (12-APR-2001) Ali Masoudi-Nejad, Kyoto University,
Graduate School of Agriculture, Laboratory of Plant Genetics,
Kilashirakawa Olive-cho, Kyoto, Kyoto 606-8507, Japan
(E-mail: masoudin@hotmail.com, Tel: 81-75-755-6145,
Fax: 81-75-753-6486)
Location/Qualifiers
1..1858
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/chromosome="1"
/map="1AS"
FEATURES
Source

```


LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
AF175312	Triticum aestivum gamma-gliadin gene, partial cds.	AF175312	AF175312.1	GI:7230477	Triticum aestivum.	
REFERENCE	Auteurs					
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
CDs						
BASE COUNT						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
Conservative						
Mismatches						
Indels						
Gaps						
Score						
DB						
Length						
927						
DB						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
151						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
210						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
211						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
212						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
213						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
214						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
215						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
216						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
217						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
218						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
219						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
220						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
221						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
222						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
223						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
224						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
225						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
226						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
227						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
228						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
229						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
230						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
231						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						
232						
CGAGAGGAATTTTTCCTCCCATATATACCAACCCCTTACACACCAACCAATTC						

[illegible]

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:46:55 ; Search time 239.66 Seconds
(without alignments)
4482.203 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 477
Sequence: 1 atgagcactacaacctg.....gcatacgagatgctctag 477

Scoring table: IDENTITY NUC
Gapox 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_101002:*

```
1: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata//geneseq//geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	477	100.0	477	21	AA245987
2	118.8	24.9	5163	19	AAV20700
3	118.8	24.9	5163	21	AAAC1849
4	118.8	24.9	5163	24	ABT04778
5	118.8	24.9	5318	19	AAV20701
6	118.8	24.9	5318	21	AAAC1848
7	118.8	24.9	5318	24	ABT04777
8	111.4	23.4	5511	21	AAA61847
9	111.4	23.4	5511	24	ABT04776

10	111.4	23.4	7334	21	AAA61846	Cryptosporidium pa
11	111.4	23.4	7334	24	ABT04775	C parvum GP900 gen
12	107.4	22.5	1107	19	AAV38816	DNA encoding a dur
13	106.2	22.3	1635	22	ABAA9946	Human breast cell
14	106.2	22.3	1635	22	ABAA7865	Human foetal liver
15	106.2	22.3	1635	22	ABAA4921	Probe #13387 for g
16	106.2	22.3	1635	22	AAK16270	Human brain expres
17	106.2	22.3	1635	22	AAK42016	Human bone marrow
18	106.2	22.3	1635	22	AAK22780	Human genome-deriv
19	106.2	22.3	1635	22	AAI22780	Human foetal liver
20	106.2	22.3	1635	22	AAI48082	Probe #12713 for g
21	106.2	22.3	1635	22	AAI08454	Probe #16768 used t
22	106.2	22.3	1635	22	ABSI6047	Human genome-deriv
23	106.2	22.3	1973	22	ABAA4805	Human breast cell
24	106.2	22.3	1973	22	ABAA5261	Human foetal liver
25	106.2	22.3	1973	22	ABAA25005	Probe #3471 for ge
26	106.2	22.3	1973	22	AAK03514	Human brain expres
27	106.2	22.3	1973	22	AAK28970	Human bone marrow
28	106.2	22.3	1973	22	AAI13556	Probe #3489 for ge
29	106.2	22.3	1973	22	AAI34918	Probe #3604 used t
30	106.2	22.3	1973	22	AAI03446	Probe #3437 used t
31	104	21.8	2198	23	ABU24090	Human genome-deriv
32	103	21.6	15720	22	AAI29504	Drosophila melanog
33	103	21.6	15720	24	ABU62358	C899P determined c
34	100.2	21.0	1075	22	ABAA48937	Human adenocarcino
35	100.2	21.0	1075	22	ABAA66852	Human breast cell
36	100.2	21.0	1075	22	ABAA33921	Human foetal liver
37	100.2	21.0	1075	22	ABAA15288	Probe #12387 for g
38	100.2	21.0	1075	22	AAK41011	Human brain expres
39	100.2	21.0	1075	22	AAK41011	Human bone marrow
40	100.2	21.0	1075	22	AAI21781	Probe #11714 for g
41	100.2	21.0	1075	22	AAI47064	Probe #7457 used t
42	100.2	21.0	1075	22	AAI07466	Human genome-deriv
43	100.2	21.0	1403	22	ABSI4983	Human breast cell
44	100.2	21.0	1403	22	ABAA3834	Human foetal liver
45	100.2	21.0	1403	22	ABAA54290	Human genome-deriv
					ABAA24047	Probe #2513 for ge

ALIGNMENTS

RESULT 1	AA245987	AA245987 standard; CDNA: 477 BP.
ID	AA245987	
XX	AA245987	
AC	AA245987	
XX	25-APR-2000	(first entry)
DT		
XX		
DE		
XX		
XX		
KW	C hordein gene; glutenin; seed storage protein; gluten; bread; pasta; noodle; breakfast cereal; snack food; cake; pastry; flour based sauce; film; coating; adhesive; building material; packaging material; grain; ss.	
KW	Hordeum vulgare.	
XX	Synthetic.	
OS		
XX		
FH		
FT		
FT	key	Location/Qualifiers
FT	CDS	1..477
FT		/tag- a
FT		/transl_except- (pos: 58..60, aa: Gln)
FT		/transl_except- (pos: 313..315, aa: Gln)
FT		/transl_except- (pos: 316..318, aa: Gln)
FT		/transl_except- (pos: 319..321, aa: Pro)
FT		/transl_except- (pos: 322..324, aa: Phe)
FT		/transl_except- (pos: 327..329, aa: Gln)
FT		/transl_except- (pos: 330..332, aa: Gln)
FT		/transl_except- (pos: 333..335, aa: Phe)
FT		/transl_except- (pos: 336..338, aa: Phe)
FT		/transl_except- (pos: 339..341, aa: Phe)
FT		/transl_except- (pos: 342..344, aa: Phe)
FT		/transl_except- (pos: 345..347, aa: Phe)
FT		/transl_except- (pos: 348..350, aa: Phe)
FT		/transl_except- (pos: 351..353, aa: Phe)
FT		/transl_except- (pos: 354..356, aa: Gln)

FT /transl_except= (pos: 355..357, aa: Glu)
 FT /transl_except= (pos: 358..360, aa: Pro)
 XX
 XX WO200002914-A1.
 XX
 PD 20-JAN-2000.
 XX
 XX 12-JUL-1999; 99WO-AU00563.
 XX
 XX 10-JUL-1998; 98AU-0004604.
 XX
 XX (CSTR) COMMONWEALTH SCI & IND RES ORG.
 PA (GOOD-) GOODMAN FIELDER LTD.
 PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
 XX
 PI Appels R, Morell M, Bekes F, Tamas L;
 XX
 DR WPI: 2000-147597/13.
 DR P-PSDB; AAY54568.
 XX
 PT Modifying glutenin or seed-storage protein for preparing foodstuffs,
 PT films, coatings, packing materials, adhesives and building materials -
 XX
 XX Disclosure; Fig 3; 76pp; English.
 XX
 XX The present sequence encodes a synthetically truncated barley C
 CC hordein protein, which is a glutenin. The protein is designated
 CC ANG-deltacyt7cy236, and is modified, according to the method of
 CC the invention. The specification describes a method for producing a
 CC modified glutenin or seed storage protein, by adding to the protein a
 CC domain that confers the ability to incorporate into gluten, or to bind
 CC a ligand or other macromolecule. The domain can be any domain that will
 CC bind ligands that may be useful in food preparation or in food
 CC composition, e.g. a domain that binds lipids or starch. The method is
 CC used for producing modified glutenins. Glutenins and seed-storage protein
 CC are useful for preparing food products such as leavened or unleavened
 CC breads, pasta, noodles, breakfast cereals, snack foods, cakes, pastries
 CC and foods containing flour based sauces. Glutenins and seed-storage
 CC protein are also useful for preparing nonfood products such as films,
 CC coatings, adhesives, building materials and packaging materials. Grain
 CC or parts of grain containing the modified glutenin and seed-storage
 CC protein is useful for preparing food products. The modified glutenins and
 CC seed-storage proteins are also useful as modifiers of food products in
 CC food industry.
 XX
 SQ Sequence 477 BP; 162 A; 186 C; 43 G; 86 T; 0 other;
 Query Match 100.0%; Score 477; DB 21; Length 477;
 Best Local Similarity 100.0%; Pred. No. 1.4e-100;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 301 CCCCAACAACCATTTCCCTGGCAACCAACACACCATTTCCCAACCCCAAGAACATTT 360
 QY 361 CAACAATAATTTTCCAGCAACCCCAACAATCATACCTGTGCAACCTCAACAGCCATTT 420
 XX
 DB 361 CAACAATAATTTTCCAGCAACCCCAACAATCATACCTGTGCAACCTCAACAGCCATTT 420
 QY 421 CCTCAACAACCTTCACACAGTCCCAACAAGCTTATCATATGATGATGCTAG 477
 XX
 DB 421 CCTCAACAACCTTCACACAGTCCCAACAAGCTTATCATATGATGATGCTAG 477
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Cryptosporidium parvum GP900 antigen open reading frame.
 XX
 KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
 KW antibody; prophylaxis; treatment; inhibition; retardation;
 KW detection; diagnosis; human; ds.
 XX
 OS Cryptosporidium parvum.
 XX
 XX
 FH Key Location/Qualifiers
 FT 1..5163
 FT CDS
 FT /*tag= a
 FT /*product= GP900 antigen
 FT misc-feature 524..1270
 FT /*tag= b
 FT /*note= "region containing NINC mutations"
 XX
 PN M09806430-A1.
 XX
 XX 19-FEB-1998.
 PD
 XX 11-AUG-1997; 97MO-US14104.
 PF
 XX 14-AUG-1996; 96US-0700651.
 PR
 XX (RBGC) UNIV CALIFORNIA.
 PA
 PI Gut J, Leech J, Nelson RC, Petersen C;
 XX
 XX WPI: 1998-159290/14.
 DR P-PSDB; AAM48299.
 DR
 XX
 PT Anti-Cryptosporidium antibody, - used to develop products for
 PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
 PT infections
 XX
 PS Claim 32; Pages 60-62; 89pp; English.
 XX
 CC The sequence is that encoding the GP900 antigen which may be used
 CC in the production of anti-Cryptosporidium antibodies. These can be
 CC used for the prophylaxis, treatment, inhibition or retardation of
 CC a Cryptosporidium infection in humans or in animals such as calves.
 CC They can also be used for the detection and diagnosis of related
 CC infections.
 CC
 XX
 SQ Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;
 Query Match 24.9%; Score 118.8; DB 19; Length 5163;
 Best Local Similarity 54.9%; Pred. No. 5.3e-18;
 Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

[illegible]

```
RESULT 3
AAA61849
ID AAA61849 standard; DNA; 5163 BP.
```

AC AAA61849

DT 28-OCT-2000 (first entry)

DE ORF encoding a portion of *Cryptosporidium parvum* NINC isolate GP900.

KW GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis; competitive inhibition; attachment; invasion; ligand binding; secreted

KW merozoite; diarrhoea; open reading frame; ORF; ds.
 XY

05 *Cryptosporidium parvum*.
XX

FH	Key	Location/Qualifiers
FT	CDS	1669 7182

```
ET      /*tag= a
ET      /partial
ET
```

```

FT      /product="Cryptosporidium parvum NINC isolate
FT      /note="No start or stop codons given in the

```

FT
XX
XX
specification"

PN 0560/1316-A.
XX

FD 00 JUN 2000.
XX

XX 12 SEP 1961 2100 00001

PR 01-JUN-1993; 93US-0071880

PR 03-APR-1995; 95US-0415751

XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
XX	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

XX

XX

DR P-PSDB; AAB11727.

PT New GP900 protein fragments and fusion proteins of *Cryptosporidium*

PT parvum, useful for detecting the presence of the parasite, and
PT diagnosing or treating cryptosporidium infections by competitive
PT inhibition of the function of GP900 -
XX
XX
Claim 16; Column 47-52; 59pp; English.
PS
XX

The invention relates to the gp900 glycoprotein of the protozoan *Cryptosporidium parvum*, DNA encoding it, gp900 fragments, and fusion proteins comprising gp900 fragments. The invention also relates to the administration of gp900 or fragments thereof to a host to elicit anti-gp900 antibody production, and to a method of cryptosporidiosis treatment or prophylaxis comprising administration of anti-gp900 antibodies to an individual. *Cryptosporidium parvum* gp900 and gp900 fragments are able to competitively inhibit sporozoite or merozoite attachment or invasion, and are also useful for the generation of anti-gp900 antibodies. The antibodies also inhibit sporozoite or merozoite attachment/invasion and additionally inhibit the binding of gp900 ligands to gp900. gp900 proteins, fragments and antibodies may therefore be used to treat or prevent cryptosporidiosis. Infection with *Cryptosporidium* is a common cause of diarrhoea in humans and causes life-threatening diarrhoea in immunocompromised persons. *Cryptosporidiosis* can be contracted from contaminated municipal water supplies (e.g., public swimming pools). It is also a cause of disease in animals, resulting in financial losses in agriculture. gp900 fragments, fusion proteins and antibodies may also be used for the diagnosis of *Cryptosporidium parvum* infections, and for the detection of the parasite in the environment. The present sequence, CC represents the open reading frame (ORF) encoding a portion of the gp900 protein of the N10c isolate of *Cryptosporidium parvum*.

SQ Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;

Query Match	24.98;	Score 118.8;	DB 21;	Length 5163;
Post Local Similarity	54.09;	Prod NO 5	30-19;	

Matches	234;	Conservative	0;	Mismatches	192;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	------	--------	----	------	----

[illegible]

RESULT 4

ID ABT04778 standard; DNA; 5163 BP.

AC ABT04778;

```

XX 27-SEP-2002 (first entry)
DT
XX
XX C parvum GP900 gene fragment SEQ ID NO: 4.
DE
XX Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KM gene; ds.
XX
XX Cryptosporidium parvum.
OS
XX WO200194631-A1.
PN
XX 13-DEC-2001.
PD
XX 14-MAY-2001; 2001WO-US15624.
PF
XX 06-JUN-2000; 2000US-0588995.
PR
XX (REGC ) UNIV CALIFORNIA.
PA
XX Petersen C, Barnes DA, Nelson RG, Gut J:
PI
XX WPI: 2002-566447/60.
DR
XX
XX Detecting Cryptosporidium in biological and environmental samples and
PT diagnosis of cryptosporidiosis involves, contacting the sample with
PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
PT
XX
XX Disclosure: Page 103-104; 157pp; English.
XX
XX The present invention relates to a method of detecting Cryptosporidium in
CC biological and environmental samples, and of diagnosing
CC cryptosporidiosis. This involves obtaining a sample and contacting it
CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC RNA, or its variant, mutant or fragment. The method is also useful for
CC detecting and identifying individual Cryptosporidium isolates based on
CC the genetic characteristics, and for diagnosis of prior or concurrent
CC cryptosporidium infection. The present sequence is a C. parvum coding
CC sequence used in the exemplification of the invention.
CC
XX
XX Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other:
SQ
Query Match 24.9%; Score 118.8; DB 24; Length 5163;
Best Local Similarity 54.9%; Pred. No. 5.3e-18;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
OY 25 CAAGAGTTGCATTCACACACATCATATCTGCCGAGCCATATCCCAAAATATTTCCG 84
DB 680 CAACAACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 739
OY 85 CTACCGCAAAACCAATTCAGTCAGAGACGCGTTTCACACACCAACCAATATTTCCG 144
DB 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 799
OY 145 TATCTACGAGAGATTTGTTCCCAATATCAATATCAAAACCCCTACACACACAA 204
DB 800 CAACTACAAACCAACAACTACAAACAAACAAACAAACCAACAACTACAAACCAAA 859
OY 205 CCATTTCCCAACACACAAACAACTCTTCTCGGCGCCCAACAAACATTTCCCTGGCA 264
DB 860 CTACAAACCAACAACTACAAACCAACAAACCAACCAACCAACCAACCAACCAAGAAAC 919
OY 265 CCACACACCAATTTCCCGAGCCCAAGAACCAATTTCCCAACCAATTTCCCTGGCA 324
DB 920 CAACAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 979
OY 325 CCACAACCAACCAATTTCCCGAGCCCAAGAACCAATTTCAACAATATATTTTCCAGACCC 384
DB 980 CAACAACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 1039
OY 385 CAACATCATACCTTGACCTCAACGCAATTTCCCAACCAACCTGCAACGATGCC 444

```

```

DB 1040 CAACAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 1099
OY 445 CAACAA 450
DB 1100 CAACAA 1105
RESULT 5
AAV20701
ID AAV20701 standard; DNA: 5318 BP.
XX
XX AAV20701:
AC
XX 17-AUG-1998 (first entry)
DT
XX Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
DE
XX Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KM antibody; prophylaxis; treatment; inhibition; retardation;
KW detection; diagnosis; human; 3' region; ds.
XX
XX Cryptosporidium parvum.
OS
XX
XX Key Location/Qualifiers
FH 1..5166
FT CDS /*tag= a
FT /product= GP900 antigen
FT misc_feature 524..1270
FT /*tag= b
FT /*note= "region containing NINC mutations"
XX
XX WO9806430-A1.
PN
XX 19-FEB-1998.
PD
XX 11-AUG-1997; 97WO-US14104.
PF
XX 14-AUG-1996; 96US-0700651.
PR
XX (REGC ) UNIV CALIFORNIA.
PA
XX
XX Gut J, Leech J, Nelson RC, Petersen C;
PI
XX WPI: 1998-159290/14.
DR
XX P-PSDB: AAW48299.
XX
XX Anti-Cryptosporidium antibody - used to develop products for
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT infections
PT
XX Claim 32: Pages 62-63; 89pp; English.
XX
XX The sequence is that encoding the GP900 antigen which may be used
CC in the production of anti-Cryptosporidium antibodies. These can be
CC used for the prophylaxis, treatment, inhibition or retardation of
CC a Cryptosporidium infection in humans or in animals such as calves.
CC They can also be used for the detection and diagnosis of related
CC infections.
CC
XX
XX Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T; 0 other:
SQ
Query Match 24.9%; Score 118.8; DB 19; Length 5318;
Best Local Similarity 54.9%; Pred. No. 5.4e-18;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
OY 25 CAAGAGTTGCATTCACCAACCAATATCTGCCGAGCCATATCCCAAAATATTTCCG 84
DB 681 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAGCACTA 740
OY 85 CTACCGCAAAACCAATTTCCAGTCAGACCAACGTTTCACACACCCCAACCAATATTTCCG 144
DB 741 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 800

```

QY 145 TATCTACGAGGAATTGTTTCCCAATATGCAATACCAACCCCTACACACAA 204
| | | | |
DB 801 CAACTACAAACCAACACTACACACACACAAACCAACCAACCAACCA 860
| | | | |
QY 205 CCATTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 264
| | | | |
DB 861 CTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 920
| | | | |
QY 265 CCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 324
| | | | |
DB 921 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 980
| | | | |
QY 325 CCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 384
| | | | |
DB 981 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1040
| | | | |
QY 385 CAACATCATACCTGTGCAACCAACCAACCAACCAACCAACCAACCAACCA 444
| | | | |
DB 1041 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1100
| | | | |
QY 445 CAACAA 450
| | | | |
DB 1101 CAACAA 1106
| | | | |

RESULT 6

AAA61848
ID AAA61848 standard: DNA; 5318 BP.

AC AAA61848;

DT 28-OCT-2000 (first entry)

DE DNA encoding a portion of Cryptosporidium parvum NINC isolate GP900.

XX GP900: NINC isolate; glycoprotein; antibody; cryptosporidiosis;

KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;

KW merozoite; diarrhoea; protozoacide; ds.

XX Cryptosporidium parvum.

XX Key Location/Qualifiers

FT CDS 1669..7182

FT /tag= a

FT /product= "Cryptosporidium parvum NINC isolate GP900"

FT /note= "No start codon given in the specification"

FT 5167..5318

FT /*tag= b

XX 3'UTR

XX US6071518-A.

XX 06-JUN-2000.

XX 12-SEP-1997; 97US-0928361.

XX 13-SEP-1996; 96US-0026062.

XX 01-JUN-1993; 93US-0071880.

XX 29-MAY-1992; 92US-0891301.

XX 03-APR-1995; 95US-0415751.

XX 14-AUG-1996; 96US-0700651.

XX (REGC) UNIV CALIFORNIA.

XX Petersen C;

XX WPI: 2000-422065/36.

XX P-PSDB; AAB11727.

XX New GP900 protein fragments and fusion proteins of Cryptosporidium

PT parvum, useful for detecting the presence of the parasite, and

PT diagnosing or treating Cryptosporidium infections by competitive

PT inhibition of the function of GP900 -

XX Claim 16; Column 41-48; 59pp; English.
PS The invention relates to the GP900 glycoprotein of the protozoan
XX Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC proteins comprising GP900 fragments. The invention also relates to the
CC administration of GP900 or fragments thereof to a host to elicit anti-
CC GP900 antibody production, and to a method of cryptosporidiosis treatment
CC or prophylaxis comprising administration of anti-GP900 antibodies to an
CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC competitively inhibit sporozoite or merozoite attachment or invasion, and
CC are also useful for the generation of anti-GP900 antibodies. The
CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC additionally inhibit the binding of GP900 ligands to GP900.
CC proteins, fragments and antibodies may therefore be used to treat or
CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC immunocompromised persons. Cryptosporidiosis can be contracted from
CC contaminated municipal water supplies (e.g., public swimming pools). It
CC is also a cause of disease in animals, resulting in financial losses in
CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC used for the diagnosis of Cryptosporidium parvum infections, and for the
CC detection of the parasite in the environment. The present sequence
CC represents genomic DNA encoding a portion of the GP900 protein of the
CC NINC isolate of Cryptosporidium parvum.
XX
XX Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other;
S0

Query Match 24.9%; Score 118.8; DB 21; Length 5318;
Best Local Similarity 54.9%; Pred. No. 5.4e-18;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 25 CAAGAGTTGCAATGCAACCAACCAATATGTCGCGCCAGCCATATCCCAAAACCAATAT 84
| | | | |
DB 680 CAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 739
| | | | |
QY 85 CTACCGCAAAACCAATTTCCAGTGCAACCGTTTTCACACACCCCAACCAATATTTCC 144
| | | | |
DB 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACCA 799
| | | | |
QY 145 TATCTACGAGGAATTGTTTCCCAATATGCAATACCAACCCCTTACACCAACAA 204
| | | | |
DB 800 CAACTACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 859
| | | | |
QY 205 CCATTCCCAACCAACCAACCAACCAACCTTTTCGCGGCCCAACCAACCAATTTCCCTGCGCA 264
| | | | |
DB 860 CTACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 919
| | | | |
QY 265 CCACAAACCAACCAATTTCCCAAGCAACCAATTTCCCAACCAACCAATTTCCCTGCGCA 324
| | | | |
DB 920 CAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 979
| | | | |
QY 325 CCACAAACCAACCAATTTCCCAAGCAACCAATTTCAACCAATATTTTCCAGCAACCC 384
| | | | |
DB 980 CAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1039
| | | | |
QY 385 CAACATCATACCTGTGCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 444
| | | | |
DB 1040 CAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1099
| | | | |
QY 445 CAACAA 450
| | | | |
DB 1100 CAACAA 1105
| | | | |

RESULT 7

ABT04777
ID ABT04777 standard: DNA; 5318 BP.

AC ABT04777;

DT 27-SEP-2002 (first entry)

XX

DE C parvum GP900 gene fragment SEQ ID NO: 3.
 XX Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
 KW gene; ds.
 XX Cryptosporidium parvum.
 OS
 PM WO200194631-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 14-MAY-2001; 2001WO-US15624.
 XX
 PR 06-JUN-2000; 2000US-0588995.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Petersen C, Barnes DA, Nelson RG, Gut J;
 XX
 DR WPI: 2002-566447/60.
 XX
 XX Detecting Cryptosporidium in biological and environmental samples and
 PT diagnosis of cryptosporidiosis involves, contacting the sample with
 PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
 PT
 XX
 PS Disclosure: Page 101-103; 157pp; English.
 XX
 XX The present invention relates to a method of detecting Cryptosporidium in
 CC biological and environmental samples, and of diagnosing
 CC cryptosporidiosis. This involves obtaining a sample and contacting it
 CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
 CC RNA, or its variant, mutant or fragment. The method is also useful for
 CC detecting and identifying individual Cryptosporidium isolates based on
 CC the genetic characteristics, and for diagnosis of prior or concurrent
 CC cryptosporidium infection. The present sequence is a C. parvum coding
 CC sequence used in the exemplification of the invention.
 XX
 SQ Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other:
 Query Match 24.9%; Score 118.8; DB 24; Length 5318;
 Best Local Similarity 54.9%; Pred. No. 5.4e-18;
 Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
 QY 25 CAAGAGTTGCAATGCACACATCATCTGCGGCGCATTCACAAACCCATAT 84
 DB 680 CAAACACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 739
 QY 85 CTACCGCAAAACCATTTCCAGTGCAGCACCGGTTTCACACACCCCAATATTTCCC 144
 DB 740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACA 799
 QY 145 TATCTACAGAGGAAATTTGTTCCCATATATCAATACCAACCCCTACACACCAACAA 204
 DB 800 CAACTACACACACAACTACAACTACAAACAAACAAACAACTACAAACCAACAA 859
 QY 205 CCATTTCCCCCAACACACACACACACCTCTGCGGCCCAACAAACATTTCCCTGGCAA 264
 DB 860 CTACAGACCAACAACTACAACTACAAACCAACCAACCAACCAACCAACCAACCAAC 919
 QY 265 CCACACACACATTTCCCGCCCAAGACCAATTTCCCAACCAACATTTCCCTGGCAA 324
 DB 920 CAAACACAAACAACTACAACTACAAACCAACCAACCAACCAACCAACCAACCA 979
 QY 325 CCACACACACATTTCCCGCCCAAGACCAATTTCCCAACCAACATTTCCCTGGCAA 384
 DB 980 CAAACACAAACAACTACAACTACAAAGAAACCAACCAACCACTACTACTACTACCA 1039
 QY 385 CAACATATATCCCTGACACCTCAACAGCCATTTTCTCAACAACTCAACAGTCCCC 444
 DB 1040 CAACACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACA 1099
 QY 445 CAACAA 450

DB 1100 CAACAA 1105
 |||||
 RESULT 8
 AAA61847
 ID AAA61847 standard; DNA; 5511 BP.
 XX
 AC AAA61847;
 XX
 XX 28-OCT-2000 (first entry)
 DE
 DE Cryptosporidium parvum Iowa isolate GP900 ORF.
 XX
 KW GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
 KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;
 KW merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.
 XX
 XX Cryptosporidium parvum.
 OS
 XX
 XX Key Location/Qualifiers
 FH 1.5511
 FT CDS /tag= a
 FT /product= "Cryptosporidium parvum Iowa isolate GP900"
 FT /note= "No stop codon given"
 XX
 PN US6071518-A.
 XX
 PD 06-JUN-2000.
 XX
 PF 12-SEP-1997; 97US-0928361.
 XX
 PR 13-SEP-1996; 96US-0026062.
 PR 01-JUN-1993; 93US-0071880.
 PR 29-MAY-1992; 92US-0891301.
 PR 03-APR-1995; 95US-0415751.
 PR 14-AUG-1996; 96US-0700651.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 PI Petersen C;
 XX
 DR WPI: 2000-422065/26.
 DR P-PSDB; AAb11726.
 XX
 XX New GP900 protein fragments and fusion proteins of Cryptosporidium
 PT parvum, useful for detecting the presence of the parasite, and
 PT diagnosing or treating Cryptosporidium infections by competitive
 PT inhibition of the function of GP900 -
 XX
 XX Claim 16; Column 37-42; 59pp; English.
 CC The invention relates to the GP900 glycoprotein of the protozoan
 CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
 CC proteins comprising GP900 fragments. The invention also relates to the
 CC administration of GP900 or fragments thereof to a host to elicit anti-
 CC GP900 antibody production, and to a method of cryptosporidiosis treatment
 CC or prophylaxis comprising administration of anti-GP900 antibodies to an
 CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
 CC competitively inhibit sporozoite or merozoite attachment or invasion, and
 CC are also useful for the generation of anti-GP900 antibodies. The
 CC antibodies also inhibit sporozoite or merozoite attachment/invasion and
 CC additionally inhibit the binding of GP900 ligands to GP900. GP900
 CC proteins, fragments and antibodies may therefore be used to treat or
 CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common
 CC cause of diarrhoea in humans and causes life-threatening diarrhoea in
 CC immunocompromised persons. Cryptosporidiosis can be contracted from
 CC contaminated municipal water supplies (e.g., public swimming pools). It
 CC is also a cause of disease in animals, resulting in financial losses in
 CC agriculture. GP900 fragments, fusion proteins and antibodies may also be
 CC used for the diagnosis of Cryptosporidium parvum infections, and for the
 CC detection of the parasite in the environment. The present sequence
 CC represents the open erading frame (ORF) encoding the GP900 protein of the

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2003, 18:55:25 ; Search time 51.7974 Seconds
(without alignments)
2824.173 Million cell updates/sec

Title: US-09-743-533-18
Perfect score: 477
Sequence: 1 atgagcgaactaaacctg.....gcataatgagatgctctag 477

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCtUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfilltest.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118.8	24.9	5163	3 US-08-700-651-1	Sequence 1, Appl
2	118.8	24.9	5163	3 US-08-928-361B-4	Sequence 4, Appl
3	118.8	24.9	5318	3 US-08-700-651-2	Sequence 2, Appl
4	118.8	24.9	5318	3 US-08-928-361B-3	Sequence 3, Appl
5	111.4	23.4	5511	3 US-08-928-361B-2	Sequence 2, Appl
6	111.4	23.4	7334	3 US-08-928-361B-1	Sequence 1, Appl
7	107.4	22.5	1107	2 US-08-991-300-1	Sequence 1, Appl
8	73.2	15.3	2455	4 US-09-103-429A-1	Sequence 1, Appl
9	73.2	15.3	2821	4 US-09-103-429A-2	Sequence 2, Appl
10	71	14.9	8442	4 US-09-272-032-6	Sequence 6, Appl
11	70.6	14.8	688	4 US-08-998-416-915	Sequence 915, App
12	70.6	14.8	10747	2 US-08-147-777-1	Sequence 1, Appl
13	70.6	14.8	10747	2 US-08-452-872-1	Sequence 1, Appl
14	70.6	14.8	10747	5 PCT-US93-03985-1	Sequence 1, Appl
15	70.6	14.8	24979	2 US-08-147-777-3	Sequence 1, Appl
16	70.6	14.8	24979	3 US-08-452-872-3	Sequence 3, Appl
17	70.6	14.8	24979	5 PCT-US93-03985-3	Sequence 3, Appl
18	68.8	14.4	198	5 PCT-US95-10668-3	Sequence 3, Appl
19	68.8	14.4	198	5 PCT-US95-10668-4	Sequence 4, Appl
20	67.2	14.1	198	5 PCT-US95-10668-2	Sequence 1, Appl
21	67.2	14.1	198	5 PCT-US95-10668-1	Sequence 1, Appl
22	66.8	14.0	1505	1 US-07-915-246-1	Sequence 1, Appl
23	60	12.6	925	3 US-08-858-003-1	Sequence 1, Appl
24	60	12.6	925	3 US-09-078-166-1	Sequence 1, Appl
25	60	12.6	925	4 US-08-997-467-1	Sequence 1, Appl
26	57.8	12.1	1086	1 US-08-415-751-47	Sequence 47, Appl
27	57.8	12.1	1608	4 US-09-292-225-20	Sequence 20, Appl

C	28	57.8	12.1	1608	4	US-09-292-225-22	Sequence 22, Appl
	29	57.8	12.1	1665	4	US-09-292-225-17	Sequence 17, Appl
C	30	57.8	12.1	1665	4	US-09-292-225-19	Sequence 19, Appl
	31	57.8	12.1	1752	4	US-09-292-225-14	Sequence 14, Appl
C	32	57.8	12.1	1752	4	US-09-292-225-16	Sequence 16, Appl
	33	57	11.9	1690	1	US-08-276-452A-24	Sequence 24, Appl
	34	57	11.9	1690	2	US-08-798-744-24	Sequence 24, Appl
	35	57	11.9	2032	4	US-09-241-581B-5	Sequence 5, Appl
	36	57	11.9	2032	5	US-08-265-428-5	Sequence 5, Appl
	37	57	11.9	2032	5	PCT-US95-07721-5	Sequence 5, Appl
	38	56.6	11.9	7785	2	US-08-276-967-1	Sequence 1, Appl
	39	53.8	11.3	2824	4	US-07-757-022B-13	Sequence 13, Appl
	40	53.8	11.3	3066	4	US-07-757-022B-83	Sequence 83, Appl
	41	53.8	11.3	3117	4	US-07-757-022B-73	Sequence 73, Appl
	42	53.8	11.3	3148	4	US-07-757-022B-57	Sequence 57, Appl
	43	53.8	11.3	3420	4	US-07-757-022B-103	Sequence 103, Appl
	44	53.8	11.3	3813	4	US-07-757-022B-43	Sequence 43, Appl
	45	53.8	11.3	3936	4	US-07-757-022B-41	Sequence 41, Appl

ALIGNMENTS

```
RESULT 1
US-08-700-651-1
: Sequence 1, Application US/08700651B
: Patent No. 6015882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEBCH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUY, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08/700,651B
: CURRENT FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415,751
: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 5163
: TYPE: DNA
: ORGANISM: Cryptosporidium parvum
: US-08-700-651-1

Query Match      24.9%; Score 118.8; DB 3; Length 5163;
Best Local Similarity 54.9%; Pred. No. 2.2e+20;
Matches 234; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY      25  CAAGAGTTGCAATTCACCAACATCATATCTGCCGACCATATTCACCAAAACCATAT 84
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      681  CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 740
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      85  CTACCCGCAAAACATTTTCAGTCAGCAACCGTTTCACACACCCCAACAATATTTCCGC 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      741  CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      145  TATCTACAGAGGAGATTTGCCCAATATCAATATCAACCCCTTACACACACACAA 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      801  CAACCTACACACACACACACACACACACACACACACACACACACACACACACACAC 860
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      205  CCATTTCACACACACACACACACCTTTCTTGCGGCCCAACACACATTTCCCTGGCAA 264
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      861  CTACACACACACACACACACACACACACACACACACACACACACACACACACAC 920
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      265  CCACACACACATTTCCCGCCCGCCAGAACCAATTTCCCAACACACATTTCCCTGGCAA 324
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      921  CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 980
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      325  CCACACACACATTTCCCGCCCGCCAGAACCAATTTCAACAAATATTTCCAGCACCC 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

DB	Sequence	Score	DB	Length
DB	981 CAACAAACAAACAACTACTACCAAGAAACCAACTACTACTACTACCA	24.9%	1040	5163
QY	385 CAACATCATACCTGTGGCACTCAACAGCCATTTCACACACTTACACAGTCCC	54.9%	444	5163
DB	1041 CAACAAACAACTACTACTACTACCAACAAACAACTACTACTACTACCA	0%	1100	5163
QY	445 CAACAA 450			
DB	1101 CAACAA 1106			

DB	Sequence	Score	DB	Length
DB	680 CAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAACTA	24.9%	739	5163
QY	85 CTACCGCAAAACCATTTCCAGTGCAGCAACCGTTTCACACACCCCAACATATTTC	54.9%	144	5163
DB	740 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA	0%	799	5163
QY	145 TATCTACCAAGAGATTGTTCCTCCATATATCAATATCAACCCCTTACACCAACAA	24.9%	204	5163
DB	800 CAACCTACCAACCAACCACTACCAACCAACCAACCAACCAACCAACCAACCA	0%	859	5163

[illegible]

RESULT 6
US-08-928-361B-1
; Sequence 1, Application US/08928361B
; Patent No. 6071518

```

1  APPLICANT: Petersen, Carolyn
2  TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
3  TITLE OF INVENTION: FOR THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
4  TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
5  TITLE OF INVENTION: SPECIES INFECTIONS
6  NUMBER OF SEQUENCES: 30
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: PETERS, VERNY, JONES & BIKSA
9  STREET: 385 Sherman Avenue, Suite 6
10 CITY: Palo Alto
11 STATE: CA
12 COUNTRY: USA
13 ZIP: 94306-1840
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patent In Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/928,361B
21 FILING DATE: 12-SEP-1997
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 60/026,062
25 FILING DATE: 13-SEP-1996
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Verny, Hana
28 REGISTRATION NUMBER: 30,518
29 REFERENCE/DOCKET NUMBER: 480.76-1(HV)
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 650-324-1677
32 TELEFAX: 650-324-1678
33 INFORMATION FOR SEQ. ID NO: 1:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 7334 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: DNA (genomic)
40 IS-08-928-361B-1

```

Query Match	23.4%	Score 111.4	DB: 3	Length 7334
Best Local Similarity	54.2%	Pred. No. 1.6e-18		
Matches 226	Conservative	0	Mismatches 191	Indels 0
			Gaps 0	
QY	34	CAATCAGCAACAACTCATATCTGCGCCAGCCATATATCAGAAACCATATCTACCCGAA	93	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	2714	CAGCAACAGCAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA	2773	
QY	94	AAACCATTTCCAGTGCAGCAACCGTTTACACACCCCAACATATTTCCCTATCTACCA	153	
		11111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	2774	CAGCCAGCAACCACTACAGCAAGAACCAACAAACAAACAAACAAACAACTACTACTA	2833	
QY	154	GAGGATTTGTTCCCATATTCATATACCAACCCCGCTACAGCAGCAACCATTTCCCG	213	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	2834	CTACAGCAACAAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTA	2895	
QY	214	CAACAGCAACAAAGCTCTTCTCTGGGCCCAACCACTTCCCTGGGCAACCAACAGA	273	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	2894	CTACTACAGCAACCAACCAACCAACCAACCACTACTACAGAAACCAACAAACAAACA	2953	
QY	274	CCATTTCCCGACCCCAAGAACCAATTTCCCAACCAATTTCCCGGCAACCAACAGA	333	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	2954	CAGCAACAGCAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA	3013	
QY	334	CCATTTCCCGACCCCAAGAACCAATTTCAACAAATATTTTCCGCAACCCCAACATCA	393	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	3014	CAACTACAGCACTACCAAGAACCAACAACTACTACTACTACTACTACTACTACTACTACTA	3073	
QY	394	TACCCGTGCAACCTTCACAGGCACTTTCTCTACCAAGCTTCAACCAAGTCCCAACAA	450	
		111 111 111 111 111 111 111 111 111 111 111 111 111 111 111		
Db	3074	CTACTACTACCAACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTA	3130	

RESULT 7
 US-08-991-300-1
 Sequence 1, Application US/08991300
 Patent No. 597325
 GENERAL INFORMATION:
 APPLICANT: D'OVIDIO, RENATO
 APPLICANT: PORCEDDU, ENRICO
 APPLICANT: MERCHITELLI, CINZIA
 APPLICANT: CARDELLI, LUISA ERCOLI
 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
 TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,300
 FILING DATE: 16-DEC-1997
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: IT MI 96/A 002663
 FILING DATE: 19-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: OBLON, NORMAN F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 2264-0201-0X
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-991-300-1

```

Query Match	22.5%	Score 107.4;	DB 2;	Length 1107;
Best Local Similarity	-56.7%;	Pred. No. 8.6e-18;		
Matches: 242;	Conservative	0;	Mismatches 176;	Indels 9;
				Gaps 2

QY	33	GCATATCCACCAACATCATATCTGCGGAGCGCATATCCAAAAACCATATTTACCGCA	92
Db	135	GCACACCAACCAACCCACTTTATGGCAACACACACACACACCATTTTGGCAGCAACA	194
QY	93	AAAACGATTTCCAGTg---CGACAAACGGTTTCACACACCGCAACATATTTCCCTATCT	149
Db	195	ACCACCATTTTACACAGCAACACACAGTTTCTACCGCAACACCATATTTTGGAGCA	254
QY	150	ACCAGAGGAAATGTGTTTCCCATATTCAAATACCAACCCCTACACACCAACCAACATTT	208
Db	255	ACAGTACCAACCATTTTGGCAGCAACACCAACACCATTTTTCACAGCAACCAACCAAGT	314
QY	210	CCCCAACAACCAACAACACTCTTCTCTGGGCCCAACACATTTCCCTGGCAACCA	268
Db	315	TTTACGGCAACAACATCATTTTGGCAGCAACATACACCATTTTACAGCAACATAC	374
QY	270	ACAACCATTTTCCCGACGCCCAAGAAACCAATTTCCCAACACATTTCCCTGGCAACCA	329
Db	375	ACCATTTTGGCAGCAACCAACAGTACTACCGCAACACACCATTTTGGCAGCAACAC	434
QY	330	ACAACCATTTTCCCGACGCCCAAGAAACCAATTT-----CAACAATTAATTTTCCGCAAC	383
Db	435	ACACCATTTTTCACAGCAACCTACCCACATTTTTCGAGCAACCAACACAGTACTACCGCA	494
QY	384	CCAACATCATACCTCTGTGCAACCTCAACAGGCATTTCTCAACACCTCAACAGTCTCC	443
Db	495	ACAACCATCATTTTTCGACACACCAACACCAACCAATTTTCACCGCAACCAACCATTTTTC	554
QY	444	CCAACCA 450	
Db	555	GCACCA 561	

RESULT 8
US-09-103-429A-1

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

1 REGISTRATION NUMBER: 34,390
2 REFERENCE/DOCKET NUMBER: BTI-39
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (607) 256-2000
5 TELEFAX: (607) 256-3628
6 INFORMATION FOR SEQ ID NO: 1:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 2455 base pairs
9 TYPE: nucleic acid
10 STRANDEDNESS: double
11 TOPOLOGY: linear
12 MOLECULE TYPE: cDNA
13 HYPOTHEetical: NO
14 ANTI-SENSE: NO
15 FRAGMENT TYPE: N-terminal
16 ORIGINAL SOURCE:
17 ORGANISM: Trichoplusia ni
18 TISSUE TYPE: peritrophic Membrane
19 IMMEDIATE SOURCE:
20 CLONE: IM14
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008

```

Query Match	15.38;	Score 73.2;	DB 4;	Length 2455;
Best Local Similarity	-49.08;	Pred. No. 2.7e-09;		
Matches 195; Conservative	0;	Mismatches 203;	Indels 0;	Gaps 0;

QY	5	CTGCGCAGACGATATCCACAAAACCATATGTACGGCAAAAACCATTTTCCAGTGCAGCAA	114
Db	291	CGCGTTTAGCCGGATGATGCACCTGGCCAGGAGACCTCCAGCTGAGACAAACCCAGGCCACACAA	350
QY	115	CCGTTTTCAGACACCCCAACCATATTTTCCCTATCTACGACGAGGAAATGTTTCCCAATAT	174
Db	351	CAACCTCAGGGCCCAACCAACCCAGGGCCCAACGACAACATCTAGTACGGCCCTCTACTCAA	410
QY	175	CAAAATACCAACCCCTTACCAACCAACCAACCAATTTCCCCCAACACCAACAACACTCTT	234
Db	411	CCACCCAGGGCCCAACCAACCAACCCAGGCCCAACCAACCAACCCAGGCCCAACCAACCA	470
QY	235	CCAGGGCCCAACCAACCAATTTCCCGTGGCAACCAACAACCAATTTTCCGACGCCCAAGAA	294
Db	471	CCGAGGCCCAATCTACACTCAGGCCCTTACTACTACACTCAGGCCCAACCAACAACAACA	530
QY	295	CCAAATTTCCCAACCAACCAATTTCCCTGGGAACCAACAACCAATTTTCCCAAGCCCCAAGAA	354
Db	531	CTGAGGGCCCTTACCAACAACCAACCCAGGCCCAACGACCAACCCAGGCCCAACTACACCC	590
QY	355	CCAATTTCAACAATTAATTTTTCAGCAACCCCAACAATCATACCTGTGCAACCTCAACAG	414
Db	591	AGGCCCAACTACCACTCAGGCCCAACATCAATCATCCAGGCTGCAACTACCCGGCGG	650
QY	415	CCATTTTCTCAACAACCTCAACCACTGCCCCCAACAAGC	452
Db	651	CAACTACCCCGGCGCAACTCCCGGCCGCGCAACTAC	688

RESULT 9
 US-09-103-429A-2
 : Sequence 2, Application US/09103429A
 : Patent No.: 6187558
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Granados, Robert R
 :
 : TITLE OF INVENTION: A NO. 6187558el Invertebrate Intestinal Mucin
 :
 : TITLE OF INVENTION: CNA and Related Products and Methods
 :
 : NUMBER OF SEQUENCES: 4
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Brown, Plimst & Michaels, P.C.
 :
 : STREET: 118 No. 6187558th Tlaga
 :
 : City: Ithaca
 :
 : STATE: NY
 :
 : COUNTRY: USA
 :
 : ZIP: 14850
 :
 : COMPUTER READABLE FORM:


```
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 915:
SEQUENCE CHARACTERISTICS:
LENGTH: 688 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1570RP
US-08-998-416-915

Query Match      14.8%  Score 70.6; DB 4; Length 688;
Best Local Similarity 49.9%; Pred. No. 8e-09;
Matches 212; Conservative 0; Mismatches 204; Indels 9; Gaps 1;

QY 34 CAATGACCAACAATCATCTGCGGAGCCATATCCAAAACCATATCTACCGCAA 93
DB 220 CAACACACAGACGACCCACGACACACACACACACACACACACACACACAC 279
QY 94 AAACCATTTCCAGTGCAGCAACGTTTCCACACACCCCAACAATATTTCCCTATCTACA 153
DB 280 CAACACACACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 339
QY 154 GAGGAATGTTTCCCATATCAAAATACCAACCCCTTACACACACACACACATTCGCC 213
DB 340 CTGTGTTACAGCTCATACGACCAACTGGCGAGGGCGGCAACCAACCAACCG 399
QY 214 CAACACACACAAACCTCTCTCTGCGCCCAACACACATTCCTGCGACACACAA 273
DB 400 CAGCAGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 459
QY 274 CCAATTTCCAGCCCAAGAACCAATTCGCCCAACACATTC-----CCCTGGCAA 324
DB 460 CAGCAGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 519
QY 325 CCACACACACATTTCCCGACCCCAAGACCAATTCACAAATATTTTCCAGCAACCC 384
DB 520 CAACACACACGCGTTCACAGCAAGTCCAGTCTCAACACACGCAAGTCCAGTCAA 579
QY 385 CAACATATATACCGTGTGACAGCCATTTCTCTCAACACGCAAGTCCAGTCCCC 444
DB 580 CAACACACACGCTTCAGCAACTTTCAGAGCTGCGCCACACACATGCAACACAG 639
QY 445 CAACA 449
DB 640 CAGCA 644

RESULT 12
US-08-147-777-1/c
; Sequence 1, Application US/0814777
; Patent No. 5914265
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: two
APPLICATION NUMBER: 07/876,289
FILING DATE: April 30, 1992
APPLICATION NUMBER: Unassigned (204/144)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/153
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10747 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-147-777-1

Query Match      14.8%  Score 70.6; DB 2; Length 10747;
Best Local Similarity 53.9%; Pred. No. 1.8e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
```

```
QY 181 CCAACCCCTTACACACCAACCAACATTCCTCCCAACACACACACACCTCTCGG 240
DB 1718 CCAACACCCCAAGCAGCAGCTCCACCAACCAACCAACCAACCAACCAAG 1659
QY 241 CCCCACACACATTCCTGCGCAACCAACCAATTTCCCGACCAAGACCAATT 300
DB 1658 CCACACACCCCAATGACCTCCACCAAGCCACACCAACCAAGCACACCAAG 1599
QY 301 CCCCACACACATTCCTGCGCAACCAACCAATTTCCCGACCAAGACCAATT 360
DB 1598 CCACACACACATTCACCAACCAAGCCACTACACGTCACCTCTGACGACACTT 1539
QY 361 CAACAAATATTTTCCAGCAACCCCAACATTCATACCTGTGCAACCTCAAGCCATT 420
DB 1538 ATGAGAGATGCTTTTACTGCGCACCAGTTTACAGACTCGGACTTCAATTCACCA 1479
QY 421 CCTCACACACCTCAACCAAGTCCCAACA 449
DB 1478 GCACCAAGCTTACCAACCAACCAACCA 1450

RESULT 13
US-08-452-872-1/c
; Sequence 1, Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
```

```
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,872
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Marburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10747 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-452-872-1

Query Match      14.8%; Score 70.6; DB 3; Length 10747;
Best Local Similarity 53.9%; Pred. No. 1.8e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 181 CCACCCCCCTACACCAACACCAATTTCCCAACACCAACCACTTTCTTCTGG 240
DB 1718 CCACCCACCCCAAGCAGCCTCCCAACCAACCAACCACTGCAAAACCAAG 1659
QY 241 CCCCAACACCAATTCCTGCGACCAACCAACCAATTTCCCAAGCCCAAGCAATT 300
DB 1658 CCACACCCCAATGCGACCTCCACCAACCAACCAACCAAGCCACCAACCAAG 1599
QY 301 CCCCAACCAATTCCTGCGACCAACCAACCAATTTCCCAAGCCCAAGCAATT 360
DB 1598 CCACACCAACCAATGCGACCAACCAACCAACCAATTCCTGCGACCACTT 1539
QY 361 CAACAATTAATTTTCAGCAACCCCAACCAATCATACCTGTCACCACTTCACGCCATT 420
DB 1538 ATGAGATGCTTTTACGCAACCAAGTTAACAGACTGCACTTCAATTCACCAACA 1479
QY 421 CCTCAACAACCTCAACCAAGTCCCAACA 449
DB 1478 GCACCAAGCTACCAACCAACCAACA 1450

RESULT 14
PCT-US93-03985-1/c
; Sequence 1, Application PC/TUS9303985
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: DEVELOPMENT OF A VECTOR TO TARGET GENE
; TITLE OF INVENTION: EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 5
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03985
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5478
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10747 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US93-03985-1

Query Match      14.8%; Score 70.6; DB 5; Length 10747;
Best Local Similarity 53.9%; Pred. No. 1.8e-08;
Matches 145; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 181 CCACCCCCCTACACCAACCAATTTCCCAACCAACCAACCACTTTCTTCTGG 240
DB 1718 CCACCCACCCCAAGCAGCCTCCCAACCAACCAACCACTGCAAAACCAAG 1659
QY 241 CCCCAACACCAATTCCTGCGACCAACCAACCAATTTCCCAAGCCCAAGCAATT 300
DB 1658 CCACACCCCAATGCGACCTCCACCAACCAACCAACCAAGCCACCAACCAAG 1599
QY 301 CCCCAACCAATTCCTGCGACCAACCAACCAATTTCCCAAGCCCAAGCAATT 360
DB 1598 CCACACCAACCAATGCGACCAACCAACCAACCAATTCCTGCGACCACTT 1539
QY 361 CAACAATTAATTTTCAGCAACCCCAACCAATCATACCTGTCACCACTTCACGCCATT 420
DB 1538 ATGAGATGCTTTTACGCAACCAAGTTAACAGACTGCACTTCAATTCACCAACA 1479
QY 421 CCTCAACAACCTCAACCAAGTCCCAACA 449
DB 1478 GCACCAAGCTACCAACCAACCAACA 1450

RESULT 15
US-08-147-777-3/c
; Sequence 3, Application US/08147777
; Patent No. 5914265
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
```

	ADDRESS:	LYON & LYON	
	STREET:	611 West Sixth Street	
	CITY:	Los Angeles	
	STATE:	California	
	COUNTRY:	U.S.A.	
	ZIP:	90017	
	COMPUTER READABLE FORM:		
	MEDIUM TYPE:	3.5" Diskette, 1.44 Mb storage	
	COMPUTER:	IBM PC compatible	
	OPERATING SYSTEM:	IBM MS-DOS (Version 5.0)	
	SOFTWARE:	WordPerfect (Version 5.1)	
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER:	US/08/147,777	
	FILING DATE:		
	CLASSIFICATION:	800	
	PRIOR APPLICATION DATA:	Including application	two
	PRIOR APPLICATION DATA:	described below:	
	APPLICATION NUMBER:	07/876,289	
	FILING DATE:	April 30, 1992	
	APPLICATION NUMBER:	Unassigned (204/144)	
	FILING DATE:	October 29, 1993	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	Wardburg, Richard J.	
	REGISTRATION NUMBER:	32,327	
	REFERENCE/DOCKET NUMBER:	204/153	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	(213) 489-1600	
	TELEFAX:	(213) 955-0440	
	TELEX:	67-3510	
	INFORMATION FOR SEQ ID NO: 3:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	24979 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	MOLECULE TYPE:	DNA (genomic)	
	US-08-147-777-3		
Query Match	14.8%	Score 70.6:	DB: 2; Length 24979;
Best Local Similarity	53.9%	Pred. No. 2.4e+08;	
Matches 145; Conservative	0;	Mismatches 124; Indels	0; Gaps 0;
QY	181 CCAACCCCCCTTACAGCACACAACCATTTCCCGCAGAACACAAACATCCTTCCTGG	240	
Db	15950 CCACGACCAGCCCAAGGCACCTCCACCAAAGCAGCACACACACACTGTGCCAATAACCCACCAAG	15891	
QY	241 CCCCAACACACATTTCCCTGGCAGCACACAACCAATTTCCCGAGGCCCAAGACCAATT	300	
Db	15890 CCACGACCCCCCAATGGCACCTCCACCAAAGCCAGCCACCAACGAACCCACCAAGAAG	15831	
QY	301 CCCCACCAACCATTTCCCGCTGGCAACCAACAACCAATTTCCCGAGGCCCAAGACCAATTT	360	
Db	15830 CCACGACACCATTAATACCAACCAACCAAGCACTGACAGCTGCATCTCTTAGCCACACTTT	15771	
QY	361 CAACCAATAATTTTCCAGCAACCCCAACCAATCATACCTTGTCGAACCTTCAGACCAATTT	420	
Db	15770 ATGGAGATGCTTTACTGCGCACAAGGTTAAACAAGATCCGACTTCCAATTCACACAGCA	15711	
QY	421 CCTCAACCAACCTCAACCAAGTCCGCCACACA	449	
Db	15710 GCACCAAGGCTACCAACCAACCAACCAACA	15682	

```
Search completed: May 25, 2003, 20:18:30
Job time : 52.7974 secs
```

